

# Understanding in Real-Time Communication

Micro-Feedback and Meaning Repair in Face-to-Face and  
Video-Mediated Intercultural Interactions



Studies in Applied Information Technology, June 2018

# Understanding in Real-Time Communication

Micro-Feedback and Meaning Repair in Face-to-Face and  
Video-Mediated Intercultural Interactions

Anna Jia Gander

Doctoral Dissertation



UNIVERSITY OF GOTHENBURG

Department of Applied Information Technology

University of Gothenburg

Gothenburg, Sweden 2018

Cover design: Catharina Jerkbrant

Cover photo and illustration: Pierre Gander & Anna Jia Gander

Understanding in Real-Time Communication  
Micro-Feedback and Meaning Repair in Face-to-Face and Video-Mediated  
Intercultural Interactions

© Anna Jia Gander 2018  
annajialugander@gmail.com

ISBN 978-91-88245-03-8  
URL <http://hdl.handle.net/2077/56223>

Doctoral thesis in Communication  
Department of Applied Information Technology  
University of Gothenburg  
SE-412 96 Gothenburg  
Sweden

Printed in Sweden, May 2018  
BrandFactory

“Everyone hears only what he understands.”

**Johann Wolfgang von Goethe**



# Understanding in Real-Time Communication

## Micro-Feedback and Meaning Repair in Face-to-Face and Video-Mediated Intercultural Interactions

### Abstract

In communication, understanding has a key position. Understanding of the understanding in real-time communication is important in both knowledge and social relevance. Previous research has pointed out that intercultural communication, video-mediated communication, and social activity with complex communication tasks and low interpersonal familiarity have more miscommunications and restrictions when it comes to understanding.

The overarching aim of this thesis is to contribute to understanding the understanding in real-time communication by empirically investigating how understanding is signalled, detected, handled, and resolved in social interactions of varying complexity in intercultural, multimodal, and video-mediated communication situations. The analytical focuses are on micro-feedback and meaning repair, using an interactional approach based on theories of social communicative activity type, meaning and implicature, contextualisation, and relevance. The thesis also aims to uncover similarities and differences in understanding between face-to-face and video-mediated communication.

Two major empirical studies have been conducted in two activity types, where the English lingua franca is spoken. Study 1 aims to investigate micro-feedback in relation to understanding issues in a spontaneous communication activity in first encounters. Based on the results from Study 1, Study 2 expands the research and aims to examine how understanding problems are coped with by acquainted interlocutors in relation to not only micro-feedback but also meaning repair in an educational activity with collaborative learning tasks.

Study 1 comprises two empirical analyses and addresses three research questions. RQ1: How are the auditory and visual modalities involved in micro-feedback expressions that are related to sufficient understanding, misunderstanding, and non-understanding? RQ2: What are the typical unimodal and multimodal micro-feedback expressions that signal sufficient understanding, misunderstanding, and non-understanding? RQ3: What specific prosodic features of vocal-verbal micro-feedback are correlated to sufficient understanding, misunderstanding, and non-understanding? The empirical material consists of eight audio- and video-recorded face-to-face interactions between four Swedish and four Chinese university students. Study 2 consists of three empirical analyses and investigates three research questions. RQ4: What are understanding and

understanding problems in social communication? RQ5: How are understanding problems detected, handled, and resolved in and through interaction? RQ6: What similarities and differences are there between face-to-face and video-mediated communication in the occurrence, detection, handling, and resolving of understanding and understanding problems? The empirical material consists of 20 audio- and video-recorded face-to-face and video-mediated interactions between ten Swedish and ten Chinese university students.

Study 1 has identified that gesture and prosody play important roles in the communication of understanding. Unimodal head movement is exclusively used to signal sufficient understanding. Eyebrow rise or frown, head forward, and gaze movement can indicate understanding problems. Smile, chuckle, and laughter associated with friendliness, politeness, uncertainty, and embarrassment can also indicate this. When nod in combination with “yeah” is associated with hesitation and uncertainty, a misunderstanding may have occurred. Sufficient understanding is found to be positively associated with both short and medium duration of micro-feedback, and non-understanding is usually communicated with a rising pitch contour. Study 2 contributes to a more nuanced classification of understanding, partial understanding. In the data, all the detected understanding problems are handled by means of meaning repair, which is either self- or other-initiated but always self-performed. The occurrence of information that is repeated, paraphrased, or responded to with unanticipated actions may indicate a misunderstanding. Video mediating technology does not seem to affect understanding, however, face-to-face communication provides better chances of detecting, handling, and resolving understanding problems. People have higher interdependency and interactivity in face-to-face communication than in video-mediated communication. Both studies show that misunderstanding does not occur as frequently as predicted in intercultural communication or video-mediated communication and is difficult for the interlocutors to detect.

Apart from enhancing the theoretical understanding of understanding in real-time communication, the empirical findings in this thesis also add to the foundation for practical design of technology enhanced education and communication, for example, online and flexible learning and digital communication, not the least in intercultural settings.

**Keywords:** understanding, micro-feedback, meaning repair, intercultural communication, face-to-face (FTF), video-mediated communication (VMC), multimodal communication, activity type, relevance, contextualisation, inference, information sharing, sense-making

# Förståelse i realtidskommunikation

Mikroåterkoppling och meningsreparation ansikte-mot-ansikte  
och i videomedierade tvärkulturella interaktioner

## Svensk sammanfattning

Kommunikation är central i mellanmännsliga aktiviteter. Kommunikation är djupt social. Den är språklig och kroppslig, kan förmedla information, kunskap och känslor, uttrycka vilja och makt, förtroende och moral. I realtid kan kommunikation ske ansikte-mot-ansikte eller — i vår samtid — i virtuella rum, medierad av digital teknologi. Förståelse har en central roll i kommunikation. Deltagarna vill uppfattas och förstås, helst på det sätt som de avsett. De vill också uppfatta och förstå vad en kommunikationspartner försöker uttrycka och förmedla. Detta är dock inte alltid lätt. Förståelseproblem uppträder regelmässigt i mellanmännslig interaktion. Det inte ovanligt att deltagare i en kommunikation förstår varandra på olika sätt eller att missförstånd uppstår, vilket kan ha konsekvenser för exempelvis mellanmännsliga relationer och möjligheterna att utföra gemensamma arbetsuppgifter.

Återkoppling fyller en viktig funktion i kommunikativa processer, bland annat för att signalera om ett budskap har uppfattats och förstås (eller inte) (McConnell, 1993; Ryan & Conover, 2004; Boud & Molloy, 2013). En särskild typ av återkoppling är s.k. mikroåterkoppling (micro-feedback) (Nivre, Allwood, & Ahlsén, 1992), som kan förstås som spontana och diskreta verbala och icke-verbala kommunikationsuttryck som används för att förmedla information om hur interaktionen kan fortsätta och hur den kommunicerade informationen har uppfattats och förstås, förutom att den förmedlar känslöstämningar och attityder. Mikroåterkoppling förmedlar enbart någon form av uppmärksamhet och förståelse, men ger inget eget innehållsligt bidrag. Mikroåterkoppling kan uttryckas i olika modaliteter (unimodalt eller multimodalt i tal och/eller gester, auditivt och/eller visuellt) eller genom olika prosodi (tonhöjd och varaktighet).

Förutom mikroåterkoppling spelar meningsreparation (meaning repair) en viktig roll för att för att upprätthålla och utveckla en kommunikativ diskurs och skapa någon form av ömsesidig förståelse genom att korrigera förståelseproblem (Clark & Wilkes-Gibbs, 1986). Det sker vanligtvis i form av att lägga till ytterligare information eller genom omformuleringar. Meningsreparationen kan initieras av den som uttrycker något (själv-initierad) eller av kommunikationspartnern (annan-initierad).

Tidigare forskning har visat på ett antal faktorer som kan bidra till att problem uppstår i kommunikation och mer specifikt problem med att deltagarna förstår varandra,

exempelvis när de har olika kulturell bakgrund, när de inte är bekanta, när kommunikationen är medierad av teknologi och när deltagarna har komplexa uppgifter att lösa gemensamt.

Denna avhandling undersöker hur människor uttrycker att de förstår och inte förstår i realtidskommunikation, framför allt genom mikroåterkoppling, vilka problem som kan uppstå samt hur de försöker lösa dessa problem, framför allt genom meningsreparation.

### **Syfte och forskningsfrågor**

Det övergripande syftet är att bidra till kunskap om förståelse i realtidskommunikation genom att empiriskt undersöka hur förståelse signaleras och problem att förstå detekteras, hanteras och löses i tvärkulturell interaktion i situationer av olika komplexitet, där engelska används som lingua franca. Det analytiska fokuset är på mikroåterkoppling och meningsreparation. Ytterligare ett syfte är att studera skillnader och likheter när det gäller förståelse och hur förståelse hanteras i ansikte-mot-ansikte-kommunikation och videomedierad kommunikation.

Två empiriska studier har utförts.

Studie 1 syftar till att undersöka mikroåterkoppling i relation till förståelse i en social och informell kommunikationsaktivitet, och består av två empiriska analyser med tre forskningsfrågor:

1. Hur är de auditiva och visuella modaliteterna involverade i mikroåterkopplingsuttryck relaterade till tillräcklig förståelse, missförstånd, och icke-förståelse?
2. Vilka är de typiska unimodala och multimodala mikroåterkopplingsuttryck som signalerar tillräcklig förståelse, missförstånd, och icke-förståelse?
3. Vilka specifika prosodiska egenskaper hos röst-verbal mikroåterkoppling är korrelerade med tillräcklig förståelse, missförstånd, och icke-förståelse?

Studie 2 syftar till att undersöka hur förståelseproblem hanteras genom mikroåterkoppling och meningsreparation. Ytterligare ett syfte var att undersöka skillnader och likheter när det gäller förståelse och hur förståelse hanteras i ansikte-mot-ansikte-kommunikation och videomedierad kommunikation. Studie 1 består av tre empiriska analyser med tre forskningsfrågor:

4. Vad är förståelse och förståelseproblem i social kommunikation?
5. Hur detekteras, hanteras och löses förståelseproblem i, och genom, interaktion?
6. Vilka likheter och skillnader finns mellan kommunikation ansikte-mot-ansikte och videomedierad kommunikation när det gäller förekomst, detektering, hantering och lösning av förståelse och förståelseproblem?

## Metod

Studie 1 undersökte mikroåterkoppling i relation till förståelse i ett första möte ansikte-mot-ansikte mellan människor som inte känner varandra sedan tidigare. Uppgiften som deltagarna i studie 1 hade var att lära känna varandra. Deltagarna var fyra svenska och fyra kinesiska universitetsstudenter. Analysens fokus ligger på fysiska aspekter av mikroåterkoppling: modalitet och prosodi i mikroåterkopplingsuttryck och innebär en kategorisering av olika typer av mikroåterkopplingsuttryck och en kvantitativ analys av frekvenser och samband mellan olika former av förståelse och sådana uttryck. Den primära teoretiska utgångspunkten tas i relevanst teori (Sperber & Wilson, 1986) och kontextualiseringsteori (Gumperz, 1982).

Uppgiften i Studie 1 var relativt enkel. Detta kan vara skäl till att det inte förekom många differentierade kategorier av förståelse eller förståelseproblem.

Baserat på resultaten från Studie 1 syftar Studie 2 till att undersöka hur förståelseproblem hanteras i en mer komplex kommunikativ miljö, där personer som redan är bekanta med varandra interagerar i en pedagogisk aktivitet med samarbetsuppgifter. Forskningsdesignen innehåller också en jämförelse mellan kommunikation ansikte-mot-ansikte och videomedierad kommunikation. Analytiskt fokus är på mikroåterkoppling och meningsreparation. Det empiriska materialet består av 20 audio- och videoinspelade interaktioner mellan tio svenska och tio kinesiska universitetsstudenter.

I Studie 2 används en annan metodologisk och teoretisk ansats än i Studie 1. Analytiskt fokus är på mikroåterkoppling, meningsreparation som dessa fenomen förekommer invävda och lokalt beroende i sekvenser av interaktion.

Begreppet förståelse undersöks genom att använda ett interaktivt tillvägagångssätt baserat på social kommunikationstyp (Wittgensteins språkspel (1958), beteende, situation och meningstyper (Allwood, 1976), aktivitetstyp (Levinson, 1979), aktivitetsbaserad kommunikationsanalys (Allwood, 2013) och kommunikativa aktivitetstyper (Linell, 2010)), konsekvenserna av avsedd betydelse och förväntad reaktion (mening och implikation (Grice, 1975)), sammanhanget av interaktion (kontextualisering (Gumperz, 1982)) och relevansdiskursen (relevanst teori (Sperber & Wilsons, 1986)).

## Resultat

Studie 1 visar att kroppsliga uttryck och prosodi spelar viktiga roller för att ge mikroåterkoppling av förståelse. I det empiriska materialet används unimodala huvudrörelser uteslutande för att signalera tillräcklig förståelse. Exempelvis höjning av ögonbryn eller rynkning av pannan, framskjutande av huvudet, blick på eller åt sidan kan indikera förståelseproblem. Leende, småskratt, och skratt associerat med vänlighet, artighet, osäkerhet och förlägenhet kan också indikera detta. Tillräcklig förståelse är positivt korrelerad med både kort och medellång längd på mikroåterkopplingen. Icke-förståelse kommuniceras vanligtvis prosodiskt med en uppåtgående tonhöjningskontur.

Studie 2 bidrar med en nyanserad klassificering av förståelse: partiell förståelse. Alla detekterade förståelseproblem hanteras i det empiriska materialet genom meningsreparation som är antingen själv- eller annan-initierad men alltid själv-genomförd. Videomedieringen tycks inte påverka förståelsen. Deltagarna uppvisar högre ömsesidigt beroende och interaktivitet i ansikte-mot-ansikte-kommunikation än i videomedierad kommunikation. I kommunikation ansikte-mot-ansikte ges bättre möjligheter att upptäcka, hantera och lösa förståelseproblem. Båda studierna i avhandlingen visar att missförstånd inte förekommer så ofta som tidigare forskning i tvärkulturell kommunikation eller videomedierad kommunikation har indikerat, och att missförstånd är svåra att upptäcka för deltagarna. Information som upprepas, parafraseras eller besvaras utan att varit förväntad, såväl som i fallet med huvudnickning och yttrandet "yeah" i kombination med tveksamhet och osäkerhet, kan innebära att ett missförstånd har skett.

### **Slutsatser**

Mikroåterkoppling signalerar inte alltid eller bestämmer förståelse i sig. Relationen mellan mikroåterkoppling och förståelse är inte så enkel, utan komplex och mångfacetterad, med många inbördes samband och eventuellt också överlappande komponenter som kanske är kända (t.ex. modalitet och prosodi, och ibland meningsreparation) och okända (t.ex. turtagande). Relationen mellan mikroåterkoppling och förståelse är starkt beroende av kontext och relevans. Å ena sidan bidrar kontexten till att förstå fenomenet mikroåterkoppling och frågan om förståelse i kommunikation. Å andra sidan bidrar mikroåterkoppling och dess prosodiska och gesturala uttryck till kontextualiseringsprocessen.

De empiriskt grundade analyserna i denna avhandling kan bidra till en mer nyanserad teoretisk och praktisk förståelse av hur människor hanterar förståelse i realtidskommunikation genom mikroåterkoppling, och hur förståelseproblem hanteras och repareras i realtidsinteraktion. Den bidrar också till att belysa skillnader mellan kommunikation ansikte-mot-ansikte och i virtuella miljöer.

Resultaten kan bidra till att utveckla riktlinjer och strategier för utformning av kommunikationstekniska applikationer, som system för tal-, gest-, och förståelse-igenkänning, grafisk visualisering och rörelsefångst och simulering. De kan också lägga grunden för praktisk utformning av utbildning och kommunikation i teknikstödda miljöer, till exempel flexibelt lärande online och digital kommunikation, inte minst i interkulturella sammanhang.

**Nyckelord:** förståelse, mikroåterkoppling, betydelsereparation, tvärkulturell kommunikation, ansikte-mot-ansikte (FTF), videomedierad kommunikation (VMC), aktivitetstyp, relevans, kontextualisering, inferens, informationsdelning, meningsskapande

# Preface and acknowledgements

When I began this research journey, I barely knew what understanding was. Then, I just wanted to develop my theoretical knowledge of language and communication and integrate them with empirical work. Because of globalisation and digitalisation, intercultural communication and communication technology are more commonly applied in social activities and individual lives. One prominent issue is that people experience miscommunication and misunderstanding more often than we think. This might be also true in mono-cultural communication. Better insights into understanding through enriched empirical analysis seem interesting and necessary in terms of both knowledge and social relevance.

This thesis would not have been possible without the support and encouragement of a number of people. First and foremost, I would like to express my sincere gratitude to my main supervisor Berner Lindström for his professional supervision and continuous support of my Ph.D. studies. He is always open-minded and encouraging, which makes my second research project possible. He is also patient and easy-going, which makes our cooperation relaxed and joyful. Thank you for believing in and supporting me. I am also very grateful to my advisor Per Linell for his inspiration and guidance as regards researching conversation and understanding. Thanks for being patient and understanding that I am still developing my knowledge and experience of the philosophy of dialogicality. Thank you for broadening my scientific horizons and prospects to combine different approaches. I am also thankful to my co-supervisor Ylva Hård af Segerstad and advisor Åsa Abelin for their specialised expert supervision and inspiring discussions. Their invaluable advice helped me a lot during this research work and writing this thesis. Especially to Ylva Hård af Segerstad, thank you for having been there for me since we first met.

Before and during my Ph.D. study, from 2003 to 2015, I had been working as a university teacher at the School of Foreign Languages and the School of Culture and Communication in Pu Tian University in China. In 2008 and 2009, I studied as a

master student at the Department of English at Uppsala University. Since 2009, I have been studying as a master student and then a doctoral student at the Department of Applied Information Technology at the University of Gothenburg. I would like to thank all the colleagues and teachers I have worked with, who helped me to build up my academic life and encouraged me to arrive here. I want to address particular thanks to Urban Nuldén and Johan Lundin who gave me enormous support and made my second research project and my entire thesis possible. Thank you for believing in me! I am also very obliged to Jan Ljungberg, Alexander Almér, Jonas Landgren, Ann-Britt Karlsson, Lotta Larsson, and Frida Sandberg. Thank you for supporting my study and work.

During the early years of my Ph.D. study, I worked at the Division of Cognition and Communication mostly with Jens Allwood and Elisabeth Ahlsén. I would like to thank them for inspiring me to research communication and micro-feedback and supervising my earlier studies. In particular, I would like to thank Jens Allwood, for opening this philosophical world to me with theoretical and methodological traditions that are classic in language and communication studies. His research philosophy has influenced me dramatically.

Thanks to Richard Hirsch, Costanza Navarretta, Patricia Poggi, Paul Hemeren, Erik Billing, and Daniel Sjölie at ICMC (International Conference on Multimodal Communication), NIC (Nordic intercultural communication), ICMI (International Conference on Multimodal Interaction), and Swecog (Swedish Cognitive Science Society) conferences for their encouragement of my research work. Special thanks to Sally Boyd for the inspiring comments and discussions at my final seminar, which guided me when finalising the thesis.

I collected the data for my first research project on face-to-face first encounters during my master thesis work at the Division of Cognition and Communication. The data were involved in the ongoing NOMCO project at that time as a comparable corpus. My second research project data on face-to-face and video-mediated task-solving collaboration were collected during my doctoral studies at the Division of Learning, Communication and IT. The data were shared with related researchers at LinCS and CogCom. I would like to thank all the participants for their participation and cooperation in the projects. Thanks to the technical support at SSKKII Research Center and at Knowledge Lab. In particular, I want to thank Mattias von Feilitzen for his practical support and effective assistance when arranging the cameras, labs, and facilities. Many thanks to Pierre Gander for his professional assistance during data collection, reliability control, statistics consultancy, and typesetting of the thesis. Also, many thanks to the transcribers, annotators, and reliability checkers Alexander Holander, Anna Valinder, Axel Olsson, Gustaf Lindblad, Karl Johan Sandberg, Max

Boholm, Pierre Gander, Tana Gegen, Tony Lorentsson, Xi Chen, Yansi Xu, and Zichen Cao (listed in alphabetical order).

The main work on the thesis was done at the Division of Learning, Communication and IT. I am thankful to my boss Johan Lundin and my dear colleagues here (too many to list) for having me in the group and allowing me independence to work. Thanks to Karin Ekman, Mattias von Feilitzen, Marie Utterberg, and Anita Grigic Magnusson for the warm discussions and company in InDesign. Many thanks to Pär Meiling, Lena Elliot, Michael Morin, Lisa Johansson, Carola Åbom, Ellen-Britta Fernell Foufa, Predrag Sucur, Marie Örtenholm Shibru, Kristina Granfeldt, Ewa Steen Stackelius, and Elisabeth Bergius for technical and administrative support.

I would also like to thank Famarz Agahi, Nataliya Berbyuk Lindström, Mikael Jensen, Alexandra Weilenmann, and Marisa Ponti for their advice and support now and then. Special thanks to Nataliya Berbyuk Lindström and Arjan Verdooren for the enjoyable collaboration in the Master in Communication program. Many thanks to my colleagues at the Division of Cognition and Communication for the inspiring seminars and informative discussions. Thank you to Alice Srugies, Ben Clarke, and Robert Lowe for your inspiring research and teaching perspectives.

Thanks to Alexander de Courcy for his language advice for both my licentiate thesis and doctoral dissertation. Thanks to Gülüzar Tuna, Kajsa Nalin, Åsa Fyrberg, Gustaf Lindblad, and Ingvar Lind for their nice company on my journey. Thanks to my friends in both China and Sweden for all their love and encouragement. I am also thankful to my other colleagues who make my time at the University of Gothenburg very enjoyable.

I would like to acknowledge Helge Ax:son Johnsons stiftelse, Stipendiefonden Viktor Rydbergs Minne, Paul och Marie Berghaus Donationsfond, Ph.D. scholarship from Chongqing Chinese Nation History Research Institute, Internationella Stipendiefond RISH (Rotary International Student House), Adlerbert Foreign Student Hospitality Fund, Travel Grant from IUC (Inter University Center), and the Department of Applied Information Technology at the University of Gothenburg, for their financial support of my research work.

Last but not least, I would like to thank my parents Xiuyan Lu and Chunxian Song who raised me with a love of science and provided me with both material and moral support. Especially in the final stage, I received tremendous help from my parents with baby-sitting and household chores so I could work and focus. Special thanks to my husband and other half, Pierre. Without you, this thesis would not have been possible today. Thank you for your love and for being there, willing to listen to me, to discuss my research, to challenge me, to give advice, and to help me with anything I needed. Thank you for everything. You are my adventure and surprise!

Many thanks to Mona, Thommy, Maria, and Thomas in Stockholm for their encouragement and support. A heartfelt thank you to Olivia for her enormous understanding and support of my choice of life and career. Thank you for your help in recruiting participants and arranging labs. Thank you, my dear, for being unique and amazing in my life. A big hug and kiss goes to my gorgeous David, for his warm and tender love calls, mamma mamma... Nu är mamma färdig! Mamma kommer!

感谢爸爸、妈妈永恒的爱与支持！  
鲁嘉 敬献

Anna Jia Gander  
Gothenburg, March 2018

# Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	The phenomenon studied.....	1
1.2	Background.....	2
1.3	Purpose and research questions.....	9
1.4	Two studies in the thesis: from Study 1 to Study 2.....	11
1.5	Organisation of the thesis .....	13
<b>2</b>	<b>Theoretical frameworks.....</b>	<b>15</b>
2.1	Communication .....	15
2.2	Communication, language, and social activity .....	17
2.3	General approaches to studying understanding .....	20
2.4	Interactional approach to studying situated understanding in real-time.....	22
2.5	Defining understanding: sense-making, meaning, and shared understanding...	25
2.6	Overt understanding in social interaction.....	26
2.7	Partially shared understanding as a basic idea.....	27
2.8	Operationalising and classifying understanding.....	28
2.9	Handling of understanding problems.....	35
2.10	Micro-feedback and understanding.....	37
2.11	Defining and operationalising micro-feedback .....	42
2.12	Summary: exploring understanding by means of micro-feedback and meaning repair.....	47
<b>3</b>	<b>Research review.....</b>	<b>49</b>
3.1	Earlier research on understanding in interaction.....	49
3.2	Research on understanding in intercultural communication.....	54
3.3	Influence of ICT on communication and understanding: VMC vs. FTF.....	56
3.4	Previous studies of micro-feedback.....	59
3.5	Research literature on prosody .....	63
3.6	Focus of the present study .....	68

<b>4</b>	<b>Method .....</b>	<b>71</b>
4.1	Method of Study 1.....	71
4.2	Method of Study 2.....	93
4.3	Ethical consideration.....	105
<b>5</b>	<b>Results of modality analysis .....</b>	<b>107</b>
5.1	Overview of the data studied.....	107
5.2	Unimodal micro-feedback.....	108
5.3	Multimodal micro-feedback.....	117
5.4	Discussion.....	124
5.5	Conclusion of Chapter 5.....	129
<b>6</b>	<b>Results of prosody analysis.....</b>	<b>133</b>
6.1	Overview of the prosodic data.....	133
6.2	Pitch and duration in relation to understanding.....	134
6.3	Pitch range related to understanding.....	140
6.4	Association between duration type and understanding.....	143
6.5	Pitch contour and understanding.....	145
6.6	Associations among pitch contour, duration type, and pitch range type.....	147
6.7	Discussion.....	153
6.8	Conclusion of Chapter 6.....	157
<b>7</b>	<b>Discussion and summary of Study 1 .....</b>	<b>159</b>
7.1	Review of the research purpose and research questions in Study 1.....	159
7.2	Summary of the empirical findings in Study 1.....	160
7.3	Contributions and implications of Study 1.....	165
7.4	Critical reflections and limitations of Study 1.....	168
7.5	Suggestions for future studies.....	171
<b>8</b>	<b>Results of reconceptualisation of understanding.....</b>	<b>173</b>
8.1	Overview of the data in Study 2.....	174
8.2	Quantifying understanding by primary means of micro-feedback.....	175
8.3	Analytical result of reconceptualisation of understanding.....	176
8.4	Discussion.....	188
8.5	Conclusion of Chapter 8.....	192
<b>9</b>	<b>Results of analysis of coping with understanding problems .....</b>	<b>195</b>
9.1	Detection of understanding problems.....	195
9.2	Handling understanding problems.....	204
9.3	Resolving understanding problems.....	207
9.4	Summary of the detection, handling, and resolving of understanding problems .....	217
9.5	Discussion.....	219
9.6	Conclusion of Chapter 9.....	221

<b>10</b>	<b>Results of comparing understanding between FTF and VMC .....</b>	<b>223</b>
10.1	Overview of the data .....	223
10.2	Descriptive statistics of the data .....	225
10.3	Basic difference between VMC and FTF in length .....	230
10.4	Occurrences of understandings in relation to communication media .....	230
10.5	How are understanding problems detected, handled, and resolved in VMC and FTF? .....	231
10.6	Discussion .....	235
10.7	Conclusion of Chapter 10 .....	238
<b>11</b>	<b>Discussion and summary of Study 2 .....</b>	<b>241</b>
11.1	Review of the research purpose and research questions in Study 2 .....	241
11.2	Summary of the results from Study 2 .....	242
11.3	Contributions and implications of Study 2 .....	247
11.4	Critical reflections and limitations of Study 2 .....	251
11.5	Suggestions for future studies .....	254
<b>12</b>	<b>General discussion and conclusion of the thesis .....</b>	<b>257</b>
12.1	Review of the research purpose, questions, data, and approach .....	257
12.2	Summary of the empirical findings .....	259
12.3	Contributions and implications of this thesis .....	263
12.4	Critical reflections and limitations of the studies .....	275
12.5	Suggestions for future studies .....	281
12.6	Concluding remarks .....	283
	<b>References .....</b>	<b>287</b>
	Appendix A: Transcription conventions .....	297
	Appendix B: Coding schemes .....	298
	Appendix C: The reading material for Study 2 .....	299
	Appendix D: General personal information questionnaire .....	305
	Appendix E: Follow-up interview .....	306
	Appendix F: Questionnaire regarding the communication experience .....	307
	Appendix G: Open discussion questions as the collaborative problem-solving tasks .....	310
	Appendix H: Consent form for participation in the project .....	311



# 1

## Introduction

This chapter introduces the studied phenomenon and its background, some key concepts, the purpose and research questions, the two studies in the thesis, and organisation of the thesis.

### 1.1 The phenomenon studied

Our human world can be viewed as a world of communication. Human communication is profoundly social, which among other things involves linguistic and cognitive information, levels of consciousness, emotions, volitions, social power, interpersonal trust, and ethics. In social communication activities, people want to communicate and want to be perceived and understood, preferably in the same or a similar way as is intended and anticipated. However, this is not easy to achieve. In reality, it is not uncommon that people are understood in many different ways, which may have various consequences for social and interpersonal communication. Understanding problems can occur at any time, and there are various types or degrees of understandings. This thesis investigates how people express understanding, what problems there might be, and how they attempt to solve the understanding problems in real-time communication.

As a key component of communication responses, *micro-feedback* has been identified as an acknowledgement of whether the communicated message has been perceived and understood. It is regarded as a signal of sense-making and information sharing in communication. Thus, in order to study how understanding is conveyed

in interaction, besides other responsive communication behaviours, the specific phenomenon of linguistic micro-feedback is focused on. Its vocal-verbal and gestural aspects are studied in this thesis. Also, prosodic aspects of micro-feedback such as pitch and duration are investigated because of their supplementary functions in communicating understanding of discourse. *Meaning repair*, as an important part of sense-making, attempts to correct the understanding problem that has occurred and been observed during the process of achieving anticipated information sharing. How meaning repair is initiated and performed during communication constitutes the process of handling understanding problems. Understanding and understanding problems are analysed here with focuses on micro-feedback and meaning repair by using an interactional approach. Considering the fact that communication technology and different cultural, linguistic, and contextual backgrounds have various impacts on understanding, face-to-face and video-mediated intercultural spontaneous communication activities of first encounters and educational task-solving collaboration will be studied. This thesis attempts to contribute to the understanding of understanding in real-time communication. It aims to provide more empirically based knowledge of how understanding is signalled, detected, handled, and resolved in intercultural interactions, where the English lingua franca is spoken, in both face-to-face and video-mediated communication activities of varying complexity with analytical focuses on micro-feedback and meaning repair.

## 1.2 Background

*Communication* can be characterised as occurring between different communicators via some communication channel or medium (e.g., sensory modality and communication technology), which has the purpose of developing and sharing information about ideas and actions and which actually has effects on such sharing irrespective of communicators' intentions (Littlejohn, Foss, & Oetzel, 2016). The general purpose of communication is to reach an understanding of both similarities and differences in knowledge.

*Language* can be viewed as *abstract objects* (i.e., strings of formal symbols and sign systems that have form and substance) and/or as *actions and activities* (i.e., sense-making interactions) (Anward & Linell, 2016). Linell (2009) has claimed that language is primarily *situated languaging* in the world, which leans more towards Saussure's (1916) *parole* (language usage) than *langue* (language system). The *langue* or *abstract objects* view is associated with formal (sentence) grammars (e.g., Chomsky, 1995). The *parole* or *actions and activities* view focuses on situated actions and

utterances in which language is primarily a resource for building utterances and contexts. Humans communicate through language in their social activity. The core of a normal communication activity is language and situation appropriateness, that is, language fits, for example, communicative needs, contextual factors such as socio-cultural and interpersonal rituals and norms, and grammaticality for comprehension, thus achieving communicative goals. When studying human communication as conversation analysis and discursive psychology reasoned methodologically, the overt interaction is the best and often the only place where cognitive processes can be observed (Hutchby & Wooffitt, 2008; Linell, 2009).

*Understanding* has a key position in human communication. In a broad sense, understanding primarily refers to the hearer's comprehending or interpreting the process of the perceived information, which is selected, organised, and evaluated (Dodd, 2005) in accordance with some assumption of *relevance* and *values* (Sperber & Wilson, 1986; Zlatev, 2009). The complex understanding process plays an important role in social signal processing and human behaviour modelling in both human-human and human-computer interactions. A large number of researchers (e.g., Allwood, 1986; Dascal & Berenstein, 1987; Weizman, 1999; Anderson, 2006; Linell, 2009; Lindwall & Lymer, 2011; Weigand, 1999, 2017) have agreed that more empirical analyses of understanding in social interaction can make significant contributions to developing the understanding of understanding in human communication.

There are already a number of different views regarding how to measure understanding in communication (Allwood, 1986; Weigand, 1999; Lynch, 2011). However, few of them have focused on how to evaluate understanding in particular in real-time *spontaneous communication*. According to Rapp and Jackson (2003, p. 236), spontaneous communication has a strong impact on mutual understanding and future cooperative activity; it is informal, not necessarily related to any specific issue, and very much dependent on face-to-face presence or some sort of simultaneous co-presence. In a similar sense, spontaneous communication in this thesis is viewed as a communication that people make up as it proceeds. The participants interdependently contribute to the joint communication, and they usually do not know in advance what they are going to say, simply because things happen to them *in the course of the communication* (Linell, 2009, p. 40).

Understanding in spontaneous communication is not easy to achieve for different reasons, for example, limitations of common knowledge and resources in *sense-*

*making*<sup>1</sup>. Human sense-making is a multi-faceted domain, and it is not easy to evaluate (Weigand, 2010; Linell, 2017). Besides, Bakhtin (1986, p. 63, p. 66, p. 170) has proposed the theory of *unfinalisability of sense-making*, suggesting that dialogic expression is *unfinalisable* and always *incomplete* and *productive* of further chains of *responses* and that *meaning* is never definitely closed and always oriented toward the future. Individuals cannot be completely understood or known. Garfinkel (1967) and Taylor (1992) have stated that we need understandings only for *current practical purposes*. Understanding one another in a communication situation is not a matter of achieving *complete and completely shared understanding* but typically achieving some “partially shared and shallow understanding” but is sufficient for us to continue with our current doings (see Linell, 2009, p. 222). Classifying and analysing understanding in human communication is methodologically problematic. Nevertheless, achieving a more effective outcome of interaction, for example, a better mutual understanding of one another is still one of the main goals in the development of communication strategies and technology. Insights into how to operationalise understanding need to be explored.

Earlier studies of understanding in interaction mostly focused on verbal aspects instead of bodily behaviours, possibly because of the tradition of *Conversation Analysis* (see discussions in Verdonik, 2010; Sayer, 2013; Goodwin, 2018). Bodily behaviours that are associated with understanding were not sufficiently studied. However, bodily aspects of communication add value to the verbal aspects in sense-making and negotiating understanding (Schul & Lamb, 1982; Stone & Posey, 2008). The visual bodily means may be just as important as the auditory verbal ones in social interaction (Borod, 1993). Thus, it requires research on evaluating understanding and its signals, both verbally and bodily.

Grice (1989) stated that understanding in communication depended on both *sentence meaning* (semantic properties of a message assigned by grammar) and *speaker’s meaning* (what the speaker intends to communicate with the *utterance*<sup>2</sup>),

---

<sup>1</sup> In this thesis, the term *sense-making* is used instead of *meaning-making*, because *sense-making* is broader in scope than *meaning-making* (see Linell, 2009; Zlatev, 2009). *Sense-making* can include all levels of awareness, consciousness, and intentionality; whereas, *meaning-making* has more to do with what is consciously meant and intended.

<sup>2</sup> In linguistic and communication studies, spoken language is normally analysed at the levels of phoneme, word, phrase, and long *utterance*. Intuitively, an utterance corresponds more or less to a clause or sentence. In spontaneous conversations, however, it is common that utterances are not clausal or sentential in form. According to the Göteborg Transcription Standard (Nivre et al., 2004), an *utterance* is a vocal contribution, that is, a continuous piece of speech beginning and ending with a clear pause. Namely, it is a sequence of words uttered by one participant bounded either by silence, or by the uninterrupted speech of another participant, or by the start/end of the communication activity. Utterances

thus there is a *relevance* in between them. *Relevance theory* was proposed by Sperber and Wilson (1986) and seeks to explain the implicit inferences in understanding communication. It states that the hearer will search for meaning in any given communication situation, and once (s)he finds a meaning that fits the expectation (s)he will stop processing. According to Vygotsky (1962), Leontiev (1981), Sarangi and Roberts (1999), and Finkbeiner, Meibauer, and Schumacher (2012), *context* bridges the explanatory gap between sentence meaning and speaker's meaning. *Context* refers to all the relevant constraints of the communicative situation, which influence language use, the discourse content, and the linguistic behaviours in communication (i.e., *contextualisation*). *Contextualisation* was first mentioned by the anthropologist Bateson (1956) and then applied specifically to language and intonation by the sociolinguist Gumperz (1982) with his *contextualisation theory*. From an interactional perspective, contextualisation primarily refers to the interdependence between speech, prosody, gesture, understanding, and discourse context (Couper-Kuhlen, 2001). For example, the prosodic aspects of speech that supplement or modify the meaning of the spoken word (Mitchell & Ross, 2013) help us to understand the speaker's meaning. *Prosody*<sup>3</sup> has a pragmatic language function and is concerned with the ways in which context contribute to meaning (Mitchell & Ross, 2013). As Nadeu and Prieto (2011) noted, assessing the levels of understanding of an utterance, "attention has to be paid to various prosodic aspects together with contextual and gestural information" (p. 841). Therefore, prosody (primarily *pitch*<sup>4</sup> and duration)

---

do not exist in written language, only their representations do. They can be represented and delineated in written language, for instance, in the transcription. Each utterance has a situated meaning (or several) of its own, interdependent with the particular matrix of contexts in which it occurs, and utterances are actions rather than behaviours (see Linell, 2009, p. 187). In this thesis, one micro-feedback item that occurs as a continuous piece of speech beginning and ending with a clear pause bounded either by silence, or by the uninterrupted speech of another participant, or by the start/end of the communication activity, is an utterance.

<sup>3</sup> Since Pollack, Rubenstein, and Horowitz (1960) and Crystal (1969), *prosody* has been basically characterised by its features with a primary focus on pitch, intensity, and duration. In this thesis, the focuses are on pitch and duration of the particular linguistic micro-feedback item. Intensity is not investigated, partly because micro-feedback usually occurs in a single word and it is not in the research interest to make comparisons between the syllables in the same micro-feedback word. Also, because the participants in this thesis project moved and gestured a lot (e.g., moved away from the microphone) in the empirical research data, it is not good enough for, or does not allow for, a study of intensity.

<sup>4</sup> According to Gulick, Gescheider, and Frisina (1989) and Smith, Patterson, Turner, Kawahara, and Irino (2005), *pitch* is a perceptual parameter and refers to the perceived tone frequency of a sound or the perceptual correlate of fundamental frequency; *fundamental frequency* (also  $F_0$ ) is an acoustical parameter and it refers to the inverse of the signal period. People measure pitch by asking people, and measure

with in particular *contextual*<sup>5</sup> information is investigated in relation to understanding in this thesis. The purpose is to examine if and how features of bodily and verbal means and the accompanying prosody can help people to acquire better insights into whether the communicated information has been understood or not.

*Intercultural communication* likely has a higher risk of misunderstanding and lack of understanding (Allwood, 2015; Lindström, 2008). As Gumperz (1982), Tannen (1990), and Samovar, Porter, and McDaniel (2012) pointed out, how people conduct their communicative behaviours in accordance with how they perceive and understand one another is strongly influenced by their own cultures. This is because *culture* is not given by nature but learned and grown up with by people. It is basically characterised by everything in terms of language, religion, cuisine, social habits, patterns of thought and behaviour, music, arts, and so forth, which are common to a particular group of people (see Smith, 2004; Eliot, 2010). When people with different cultural and linguistic backgrounds bring their local knowledge together into a joint communication activity, it is very likely that they have more problems and difficulties in understanding each other than when they communicate with people who have the same relevant backgrounds. In this thesis, how *Swedish and Chinese speakers*<sup>6</sup> cope with understanding in interactions is studied. Swedish and Chinese cultures are chosen, because of significant physical, regional, linguistic, and social differences and increasing global cooperation.

Vygotsky (1962), Garfinkel (1967), Goffman (1971), Rommetveit (1974), Grice (1975), Schegloff (1987), Allwood (1993), and Linell (2010) at different times have agreed that social structure is constituted in social practice, especially communicative practice. Social communicative activity type plays a large role in communication and understanding. Social communicative activity type, on the one hand, constrains what will be taken up as allowable contributions by the participants and, on the other, helps the participants to determine what kinds of inferences will be made from what has been said (Grice, 1975; Levinson, 1979). “Because these activity-specific rules of inference are more culturally specific than other sorts, they are likely to play a large role in cross-cultural or interethnic miscommunications” (Levinson, 1979, p. 393).

---

fundamental frequency or  $F_0$  by asking the computer. Under normal circumstances, *fundamental frequency* ( $F_0$ ) of an acoustic voice signal is the *primary determinant* of the perceived pitch of the voice (Kreiman & Sidtis, 2011, p. 26).

<sup>5</sup> *Contextual* primarily means being context dependent and/or interdependent.

<sup>6</sup> A *Swedish/Chinese speaker* refers to a person, who was born in Sweden/China, speaking Swedish/Chinese as his/her first language, having Swedish/Chinese as the dominant culture, and has been primarily living in Sweden/China before the age of 18.

Equally important, Navarretta and Paggio (2013) and Campbell (2007) have stressed that among other factors, such as the physical setting, the number of participants, and the topic discussed, the degree of familiarity influences a lot the use of micro-feedback and mutual understanding. Because of the different individual communication presuppositions and expectations as well as various limitations of common knowledge and resources in sense-making (Linell, 2009), the unacquainted people who have mutually distinct and unknown personal and professional experiences differ from the acquainted people, when it comes to achieving shared understanding in communication (Maynard & Zimmerman, 1984; Kiesler & Sproull, 1992; Svennevig, 1999). Accordingly, communication activities of varying complexity, *strangers' first encounter* and *acquaintances' task-solving collaboration*<sup>7</sup>, are focused on in this thesis.

Contextual and technological influences on communication and understanding have been recognised in different communicative activities. Understanding is more salient in communication of complex tasks than that of simple ones. For example, because successful learning cooperation requires intensive high quality discourse exchange (i.e., not only talk or interact but also negotiate and (re)structure knowledge and produce learning outcome), understanding is more crucial in a learning cooperation context (Lindwall, Lymer, Lindström, & Bernhard, 2005; Sins, Savelsbergh, van Joolingen, & van Hout-Wolters, 2011) than during, for instance, a coffee break chat. Also, it has been found that participants have more difficulties in a joint problem-solving task when they use technology-mediated communication instead of face-to-face communication (Hancock & Dunham, 2001). Compared to face-to-face, individuals in video-mediated communication are more unaware of each other's non-verbal behaviours and they are more constrained when it comes to achieving understanding (e.g., Olson & Olson, 2000). Some research has found that mediating technology has little effect on interaction and understanding (e.g., Anderson, 2006). Thus, the issue of whether understanding varies between face-to-face and video-mediated communication situations needs to be researched.

McConnell (1993), Ryan and Conover (2004), and Boud and Molloy (2013) among others have identified *feedback* not only as an acknowledgement of whether the information has been perceived and understood but also as an important part of

---

<sup>7</sup> In this thesis, *strangers* and *unacquainted* both refer to people who do not have any previous acquaintance with or knowledge about one another (but they have probably various assumptions). *First encounters* not only refer to the first meetings or first joint experiences between strangers or unknown things, but also emphasise the interaction and interactivity between them in this thesis. Also, in the present study, *acquaintances* are people who have had previous contact or an association with or knowledge about one another and may not be close friends.

the communication exchange process that eventually leads to meaning and understanding sharing. Just as Nadeu and Prieto (2011) suggested, the very response feedback to the presented information can be perceived and understood as conveying more or less information on understanding depending on how much information about the context the hearers possess. “Feedback responses provided by recipients are one of the explicit marks for achieving understanding” (Bertrand & Goujon, 2017, p. 53). In this thesis, how to interpret understanding through feedback and context of relevance is investigated in the communication activities.

Although primarily derived from Wiener’s (1948) cybernetic notion, *feedback* as a concept and a terminology has been widely used across various disciplines. Even in the same field of linguistics and communication, it is used in different ways. For example, *feedback* is commonly used to refer to the communication responses as full contributions to the discourse (Wood, 2011, p. 85), for example, *how are you? I’m fine thank you...* and *what’s your name please? My name is Julia...*, or the comments and evaluations which are often focused on in applied linguistics or sociolinguistics and mainly made up of comprehensive and expanded responsive expressions (Mahboob & Knight, 2010, p. 31), for example, *well done boy, very good, you’ve made big progress; there are only two misspellings here...* and *I appreciate your work and you have some good points about... but ... is not clear...*, or some particular linguistic device that has certain communicative functions (Nivre, Allwood, & Ahlsén, 1992, p. 1), for instance, *yeah yeah yeah* or head nod that signals “I hear and understand what you have just said”, or something similar. In order to reduce the terminological ambiguity in the field of linguistics and communication, one distinct feedback phenomenon, named *micro-feedback*, is specified in this thesis. Most language and communication researchers seem to agree that it primarily refers to “the unobtrusive verbal and nonverbal communicative expressions that are used to give and elicit information” (Nivre et al., 1992, p. 1) about the continuation of the interaction, the perception and understanding of the information communicated, as well as the attitudinal and emotional reactions to the perceived and understood information (see Nivre et al., 1992). The attribute “unobtrusive” means that with *micro-feedback* expressions the interlocutor who is using them does not claim to make his/her own contribution to the main discourse, and that the interlocutor does not intend to take the turn but only signals some kind of attention, perception, and understanding without making any substantial contribution to the interaction. That is, *micro-feedback* is spontaneous and unobtrusive vocal-verbal and gestural expressions such as *yeah, okay*, head nod, and smile, which depend heavily on context rather than referential or semantic meaning and express positive and negative evaluative opinions.

A number of previous studies of this feedback phenomenon have been conducted, although with different purposes such as identifying it as mostly produced by listeners<sup>8</sup> in interactions (Goodwin, 1981), describing various ways of producing it (Lu & Allwood, 2011), exploring the complementary information that one of its *modalities*<sup>9</sup> (e.g., auditory and visual) can provide to another modality (Cerrato & Skhiri, 2003), analysing the affective and interactive aspects of it (Navarretta, Paggio, & Jokinen, 2008), and analysing the types and functions of it in human-human and human-computer interactions (Paggio & Navarretta, 2013). However, little has been done to identify how micro-feedback can be studied as regards the features of its modality and prosody. Likewise, little work has been done in investigating the relations between micro-feedback and understanding, for example, in which way and to what extent which type of micro-feedback is in relation to what kind of understanding and whether micro-feedback can be used as a signal to identify and analyse understanding and understanding problems. The current study addresses these questions.

Besides micro-feedback, *meaning repair* plays an important role in forming and reforming understanding and developing discourse exchange (e.g., Clark & Wilkes-Gibbs, 1986). It usually takes place in the form of adding further information or changing the original information into a new information. In line with this, *meaning repair* in this thesis refers to a communicative action that attempts to correct the understanding problem that has occurred and been observed during the process of achieving anticipated information sharing. The meaning repair process continues until a shared understanding is formed and accepted for the current communication purposes. Since meaning repair is common in the understanding process, it will be researched when studying understanding in real-time communication.

### 1.3 Purpose and research questions

The overarching aim of this thesis is to contribute to understanding the understanding in real-time communication by empirically investigating how understanding is signalled, detected, handled, and resolved in social interactions of varying complexity in intercultural, multimodal, and video-mediated communication situations. The

---

<sup>8</sup> In this thesis, a *listener* or *listening person* (or *hearer*) only refers to the current status of the interlocutor in particular regarding the relation between the current utterance and the subsequent one. A *speaker* or *speaking person* is the opposite, which refers to the current status of the interlocutor with respect to the earlier utterance.

<sup>9</sup> *Modality* is used to refer to the sensory modality (see Section 2.11).

analytical focuses are on micro-feedback and meaning repair, applying an interactional approach. The thesis also aims to uncover similarities and differences in understanding between face-to-face (FTF) and video-mediated communication (VMC). Studying understanding in particular in relation to micro-feedback necessitates conceptual analyses of the concepts of understanding and micro-feedback.

Two major empirical studies have been conducted in two activity types, where the English lingua franca is spoken. Study 1 aims to investigate micro-feedback in relation to understanding issues in a spontaneous communication activity in first encounters. Based on the results from Study 1, Study 2 expands the research and aims to examine how understanding problems are coped with by acquainted interlocutors in relation to not only micro-feedback but also other responsive interactions, primarily, meaning repair in an educational activity with collaborative learning tasks.

More specifically, Study 1 comprises two empirical analyses and addresses three research questions (RQ1–3).

(1) Analysis of modality:

RQ1: How are the auditory and visual modalities involved in micro-feedback expressions that are related to sufficient understanding, misunderstanding, and non-understanding?

RQ2: What are the typical unimodal and multimodal micro-feedback expressions that signal sufficient understanding, misunderstanding, and non-understanding?

(2) Analysis of prosody:

RQ3: What specific prosodic features of vocal-verbal micro-feedback are correlated to sufficient understanding, misunderstanding, and non-understanding?

Study 2 consists of three empirical analyses and addresses a further three research questions (RQ4–6).

(1) Reconceptualisation of understanding:

RQ4: What are understanding and understanding problems in social communication?

(2) Analysis of coping with understanding problems:

RQ5: How are understanding problems detected, handled, and resolved in and through interaction?

(3) Comparing understanding between FTF and VMC:

RQ6: What similarities and differences are there between face-to-face and video-mediated communication in the occurrence, detection, handling, and resolving of understanding and understanding problems?

More nuanced insights into how understanding and understanding problems are coped with in real-time communication will be presented.

## **1.4 Two studies in the thesis: from Study 1 to Study 2**

This thesis consists of two empirical parts. Study 2 is an expansion and development of Study 1.

Study 1 focuses on investigating micro-feedback as a signal of understanding based on spontaneous intercultural face-to-face first encounters. Features of the modality and prosody of micro-feedback in relation to different forms of understanding are investigated in detail. It is likely because the communication activity task of getting acquainted with one another is relatively simple, the unacquainted but friendly and kind participants easily became acquainted and did not generate many differentiated categories of understanding and understanding problems.

As Brown and Yule (1983) have argued, in an interactional discourse such as encounters the interlocutors are primarily concerned with maintaining friendly social relations rather than pursuing mutual understandings; whereas, in a transactional discourse (see a more detailed explanation in the original article) such as learning and task-solving activities, accurate transfer of information and sufficiently shared understanding are the main concern. In order to conduct a further study of understanding, a more complex educational task was used in Study 2.

Also, although face-to-face communication is historically seen as the basis of a theory of language, the basis of all human language behaviour, and the standard communication situation (Clark, 1996), there could be a possibility that face-to-face communication produces certain numbers and types of understanding problems. Technology-mediated communication can provide complementary perspectives. In particular in this digital era, there is a growing interest in technology enhanced education and communication, for example, online and flexible learning and digital communication. Accordingly, Study 2 in this thesis was designed to enrich the research data across varying communication media. Based on the analyses and results found in Study 1, Study 2 expands the data from face-to-face only to face-to-face and video-mediated communication.

Study 2 focuses on examining how understanding problems are coped in a more complex communicative setting where acquaintances are given educational tasks to

solve in face-to-face and video-mediated collaboration. With a more complex design than Study 1, more understanding problems were anticipated from Study 2. Building on the new data, further investigations will contribute to a more nuanced comprehension of how understanding is coped with in social interaction.

The analytical focuses in the thesis have also been expanded and developed from Study 1 to Study 2. In Study 1, the phenomenon of micro-feedback was explored based on relevance theory and contextualisation theory, with a focus on context dependency. The unit of analysis was micro-feedback, and the main classifications were based on physical properties of micro-feedback expressions, even though the analysis was not completely lacking in interactional considerations. Empirical results show that participants draw other resources, for example, meaning repair, in addition to micro-feedback in coping with understanding and understanding problems. Therefore, in Study 2, a different and evolved approach is used with analytical focuses on micro-feedback and meaning repair. The concept of understanding is examined by using an interactional approach based on theories of social communicative activity type, meaning and implicature, contextualisation, and relevance. The analyses are focused on interactions (i.e., interactional sequences), in which utterances are assumed to be sequentially intertwined, that is, interdependent within the interactional sequence. Based on the empirical data, the study of understanding in real-time communication also carries out a more extended conceptual analysis of understanding compared to Study 1.

The analyses in this thesis were performed from an analytical perspective. The findings and discussions expand on previous research on understanding in real-time communication by adding an extensive literature review, complementary research methods, empirical material from different activity types, and multiple analytical focuses.

This thesis contributes to a more nuanced theoretical and practical understanding of how people cope with understanding in real-time communication, in particular with more enriched empirically based knowledge. The results can serve as a foundation for the development and design of communication technology applications, such as systems for speech, gesture, and understanding recognition, online and flexible learning, digital communication, and video-related innovations and applications. Self-recognition and awareness of being sufficiently understood, partially understood, misunderstood, and not understood can be enhanced in intercultural communication practice.

## 1.5 Organisation of the thesis

The thesis consists of twelve chapters. Chapter 1 presents the introduction of the phenomenon studied, the background, the purpose and research questions, and the two studies in the thesis. In Chapter 2, theoretical frameworks and the rationale of the approaches to analysing understanding and micro-feedback are presented. Chapter 3 presents the research review of understanding and micro-feedback. What has or has not been done in earlier research to identify and operationalise micro-feedback and understanding is discussed. Chapter 4 deals with the research method of the thesis, that is, how Study 1 and Study 2 were designed and how the research data were collected and processed.

This is followed by Study 1. Two empirical analyses and the results are presented. In Chapter 5, micro-feedback is investigated in detail with respect to its modality. What the typical unimodal and multimodal micro-feedback expressions are that signal sufficient understanding, misunderstanding, and non-understanding are identified. Chapter 6 primarily investigates the prosody of micro-feedback in relation to understanding. The specific prosodic features (i.e., pitch and duration) of the vocal-verbal micro-feedback that are correlated to sufficient understanding, misunderstanding, and non-understanding are identified. Chapter 7 presents the discussion and summary of Study 1. Research questions (RQ<sub>1-3</sub>) and empirical findings of the relation between micro-feedback and understanding are summarised.

Study 2 will be presented as follows. Chapter 8 presents a theoretical analysis of reconceptualisation of understanding in interaction with respect to what understandings are. Chapter 9 contains an empirical investigation of how understanding problems are detected, handled, and resolved through micro-feedback and meaning repair in and through communication and how participants construct sense-making and information sharing. Chapter 10 presents a comparative study on the occurrence, detection, handling, and resolving of understanding and understanding problems between FTF and VMC. A discussion and summary of Study 2 is presented in Chapter 11. Research questions (RQ<sub>4-6</sub>), empirical findings, and theoretical assumptions of understanding and understanding problems are summarised.

Finally, Chapter 12 presents the discussion and conclusion of the entire thesis. The empirical findings regarding understanding in real-time communication, contributions, theoretical and practical implications, and critical reflections and limitations of the studies are summarised, and suggestions for future work on understanding in communication are discussed.

The transcription conventions and coding schemes used in the thesis are presented in Appendix A and Appendix B, respectively. The reading material, the general personal information questionnaire, the follow-up interview, the questionnaire regarding the communication experience, the open discussion questions as the collaborative problem-solving tasks, and the consent form for participation in the project are presented in Appendix C, D, E, F, G, and H, respectively.

# 2

## Theoretical frameworks

In this chapter, theories of communication, language, and social activity are presented in Sections 2.1 and 2.2. Theories of and approaches to studying understanding and understanding problems in situated social interaction are presented in Sections 2.3 to 2.7. Section 2.8 presents the operationalisation and classification of understanding in social communication. Meaning repair in handling understanding problems is presented in Section 2.9. In Section 2.10, theories of how understanding is manifested through the responsive behaviour of micro-feedback are discussed. The concept of micro-feedback and the theories and approaches to studying micro-feedback are addressed in Section 2.11. Section 2.12 presents a summary of how understanding is studied with micro-feedback and meaning repair as focuses in this thesis.

### 2.1 Communication

Communication as a term stands for somewhat different (more or less broadly or narrowly defined) phenomena in different contexts. In this section I will briefly overview some approaches and definitions.

Communication takes place in interactions<sup>10</sup> between two or more parties (participants, interlocutors), in which at least one or more of the participants (normally

---

<sup>10</sup> Based on Linell's *situated interaction* (see 2009, p. 98) which involves both the concrete interactive situation (e.g., who, when, and where) and the continuously changing discourse content, *interaction* in this thesis, primarily refers to the situated mutual influencing, interactive, interdependent, and jointly coordinated communicative action in a certain situation between (at least two) individual *interlocutors*. The interlocutors affect each other and jointly coordinate, develop, and actively co-construct the interaction.

human beings, but occasionally higher animals and advanced artificial systems) are involved in activities of sense-making, producing, or understanding messages related to self (selves), other(s), and/or environments. There are many alternative definitions (Krippendorff, 1993). A classical definition is that of *information transfer*; for example, Shannon and Weaver (1949) used it when dealing with “objective” (person-independent) information, and the transfer capacities of technical systems. This did not account for knowledge and understandings of human beings. This definition also gave priority (sometimes even full priority) to the sender (“the communicator”), while an analysis of purposive, meaningful interaction would have to consider both (or all) participants (interlocutors).

So, what about defining communication as *information exchange*? This would be closer to a workable notion to be used in interaction analysis (including the analysis of verbal interaction). But if it is confined to cognitive information, it is still not quite what we need. We need to also include different levels of consciousness. Allwood (1976) provided an extensive, normative, and conceptual analysis of different forms of communication. Different levels of awareness or consciousness are possible on both the production side and the reception (understanding) side (even though they are probably continuous scales). For example, recipients often provide (micro-) feedback on the basis solely of *shallow* understandings (close to what Allwood called “pure apprehension”, pp. 90–91).

According to Allwood, participants indulge in intentional and purposeful cooperation (at some level). Information can be indicated (information expressed without communicative intention), displayed (intention to show), and signalled (intention that recipient recognises intention to show as in the case of symbolic communication via language, provided that it is known and understood by the participants). Both speakers and recipients can use information and understand it at different levels of consciousness. On the least aware and intentional level (the indicative level), both speaker and listener are not aware of that communication and influence occurs “automatically”, which leads to responsive behavioural reactions that are hard to control (see Allwood, 2002), for instance, a head nod and small words *yeah okay*. On the middle level (the display level), both speaker and listener are more in control or aware of the communication behaviour that has been initiated as indicated and automatic and then gradually becomes intentionally displayed, such as an eyebrow rise, a straight gaze at (the speaker), and a surprising word such as *really*. The “highest” form of communication at the most aware and intentional level (the signal level), with both speaker and listener operating with—understanding—information as signalled (symbolically loaded), is “full-blown” communication. For example, an okay

gestural signal<sup>11</sup>, an eliciting question *what do you mean by... sorry I don't understand* seeking further explanation and clarification, and an assertive declaration *yes definitely because...* expressing the confirmative attitude and the agreed evaluative opinion are all at this level.

Furthermore, not only cognitive aspects and intentional levels need to be attended to, but also emotions, volitions, social power, social and interpersonal relation, ethics, and so forth that are involved in and relevant for the communication activity. In addition, participants must trust each other (at some level), the speaker must not inflict pain or injury on the recipient, and (s)he must leave leeway (some freedom) to the recipient to contribute.

Allwood's thesis (1976) (Note: in later texts, Allwood has modified or developed his conceptual apparatus) was, however, a conceptual analysis rather than an empirical analysis of interaction in real-time (as, for example, the present dissertation).

Many scholars have assumed that the product, or at least the goal, of successful communication is shared knowledge or information (cf. "common ground" in Clark, 1996) or shared understanding (Schegloff, 1991) of the given topic (cf. Lat. *communicatio* from *communis* 'common, shared'). However, this would be realistic only as far as simple pieces of information are involved, for example, information and dialogue about basic personal information such as sex, age, birth-place, domicile, and so forth (as in the first encounter's communication activity of Study 1). Completely shared understanding is therefore seldom at hand. The typical case is instead *partially shared* understandings, which is, or has to be, *sufficient for current situated purposes* (cf. Garfinkel, 1967).

Finally, note that the purpose of a communicative project in a joint dialogue need not be shared understanding with the other; it may be about identifying differences in knowledge or understandings that the participants hold or come to entertain.

Several of the themes introduced here will be further developed in subsequent sections and elsewhere in the dissertation.

## 2.2 Communication, language, and social activity

Communication, language, and social activity and their interrelations have been of great interest in academia. Saussure (1916) has assumed that language is an example

---

<sup>11</sup> The hand gesture performed by forming a circle (the O) with the thumb and index finger and holding the other fingers straight or relaxed in the air is commonly used in bodily communication. In many parts of the world (e.g., US and most of Asia and Europe), it is synonymous with the word OK, denoting approval, agreement, or that all is well. In other contexts or cultures (e.g., France, Belgium, Japan, and Arabic countries), this gesture may have different meanings or connotations, including negative or offensive ones (e.g., Brazil, Russia, Germany) (cf. de Mooij, 2013, e.g., p. 26).

of an ontologically irreducible social phenomenon. Similarly, Durkheim (1895) exemplifies social ontological holism in that sociology should study phenomena attributed to society at large rather than being limited to the specific actions of individuals. By contrast, Popper (1957) and Homans (1961) exemplify reductionist social behaviourism and methodological individualism by stating that any generalisation about human social life is derivable from the basic principles that are true of individuals instead of groups or cultures. This is also true of American *social* psychology (Allport, 1924; Farr, 1996).

When it comes to discussing the nature of linguistic communication, the philosopher Wittgenstein (1958) developed the theory of language games that the meaning of linguistic expressions is given by their use in language games. In a similar vein, Austin (1962) and Searle (1969) proposed the theory of speech acts, that is, to speak is to act. Grice (1975) posited the theory of meaning and implicature, that to speak is to act rationally. Levinson (1979) proposed the theory of activity type, that language is expected to play roles and functions within specific activity types, which is an idea similar to Allwood's (1976) theory of *behaviour, situation, and meaning types* and his *activity-based communication analysis (ACA)* (Allwood, 2013) as well as Linell's (2010) *theory of communicative activity type (CAT)*.

With the theory of language games, Wittgenstein (1958) rejected Saussure's (1916) idea that language is separate from reality, and proposed that the entirety of language does not consist of the forms of language but also of the actions into which language is woven. A concept does not need to be clearly defined to be meaningful, instead it acquires various meanings in different actions. Speaking of a language is part of an activity that gives language its meaning. Saying something in a language is analogous to making a move in a game (Wittgenstein, 1958). Words have meanings depending on the uses made of them. Understanding language and communication is "by implication having a grasp of the meaning of utterances, involves knowing the nature of the activity" (cf. Levinson, 1979, p. 365).

The theory of speech acts can be traced back to Austin's (1962) performative utterances and locutionary, illocutionary, and perlocutionary acts. Kempson (1977, p. 51) has summarised the three interrelated speech acts as "a speaker utters sentences with a particular meaning (locutionary act), and with a particular force (illocutionary act), in order to achieve a certain effect on the hearer (perlocutionary act)". Austin's idea was further developed by Searle (1969; 1979) as the contemporary theory of speech acts. The theory of speech acts is a subfield of pragmatics concerned with the ways in which words can be used not only to present information but also to carry out actions. Speech acts are commonly taken to include such acts as promising, ordering, greeting, warning, inviting, and congratulating (see Martínez-Flor & Usó-

Juan, 2010; Barron, 2003). A “speech act” is also an utterance that has a performative function in language and communication. Schegloff (1992, p. 1339) has criticised Searle's speech-act theory for “cast[ing] action as atomistic, individualistic, atemporal, asequential, and asocial”.

Regarding the theory of meaning and implicature, Grice (1975) proposed to take logic into account for conversation and look at the utterer's meaning and intentions together with sentence meaning and word meaning in language communication. His approaches later came to be called “intention-based semantics” and “activity-specific rules of inference”, which are popular in language pragmatics (cf. Allwood, 1993). The focal point is to explain language meaning based on the implicature or inference of the speaker's intentions. To put it differently, to speak is to perform a central and direct speech act, and to implicate is to perform a non-central and indirect speech act.

Regarding the relation between communication, language, and social activity, first, psychologist Bühler (1934) has pointed out that language must take into account its social interactive functions, which has been also stressed by Vygotsky (1962) and Rommetveit (1974). Second, the sociologist and pragmatist Mead (1934) also focused on the pragmatic and symbolic aspects of interaction between people and the world and their dynamic relations. Mead (1982) has said that “individual mind can exist only in relation to other minds with shared meanings” (p. 5) and physical objects are social objects with also rich meanings. Such ideas are followed up in Linell's (2009) dialogism. Third, a large number of researchers (e.g., Garfinkel, 1967; Goffman, 1971; Schegloff, 1987; Grice, 1975) have more or less agreed that social structure is constituted in social practice, especially communicative practice (cf. Allwood, 1993, 2000). This corresponds to the linguist Firth's (1957) claim that language can be understood only in a social and anthropological context and Levinson's (1979) activity type theory that language is expected to play roles and functions within specific kinds of social activity.

As regards the activity type theory, Levinson (1979), Goffman (1974), and Gumperz (1982) have all pointed out that activity type plays an important role in understanding communication. According to their work, an activity type refers to any culturally recognised activity in which the participants' actions are goal-defined, socially constituted, and bounded with constraints on participation and contribution.

Knowledge of the conversational activity entails expectations about possible goals or outcomes for the interaction, about what information is salient and how it is likely to be signaled, about relevant aspects of interpersonal relations, and about what will count as normal behavior. (Gumperz, 1982, p. 101)

Social activity type plays a central role in language usage. On the one hand, social activity type constrains what will count as allowable contributions to the communication. On the other hand, social activity type helps to determine what kinds of inferences will be made from what is communicated. That is, communication is a creative sense-making activity; with a knowledge of the activity type, communication involves conforming the abstract rules of the activity and creating the joint interactions in line with activity and participant expectations; both are done with a low consciousness level. Levinson (1979) has contributed to the distinction between speech acts and speech activities, both of which fall under the theory of language games, which has given prominence to “the fact that the speaking of language is part of an activity or a form of life” (Wittgenstein, 1958, p. 23). This relation between what is said and what is done is of central importance to the study of understanding, and this makes it possible to research human interaction.

### 2.3 General approaches to studying understanding

Language is a system of correlating meaning and expression, and the general purpose of communication is to come to an understanding. The assumed difficulty in explaining what understanding is, shared by philosophers, psychologists, educators, and linguists (see Helmstad, 1999), has been approached in one way or another in the field of philosophy, psychology, cognition, and pedagogy.

From a philosophical perspective, some philosophers of language believe that understanding is inextricably linked to believing (Romero-Trillo, 2010). For instance, Edwards (1967) associated understanding with interpretation, which adheres more to *the hermeneutical tradition*<sup>12</sup> than *the epistemological tradition*<sup>13</sup> in trying to find out what understanding really means in philosophy. In this approach, explicit

---

<sup>12</sup> According to Palmer (1969), the term *hermeneutics* covers both the first order art and the second order theory of understanding and interpretation of linguistic and non-linguistic expressions. As a theory of interpretation, the *hermeneutic tradition* stretches all the way back to ancient Greek philosophy. During the Middle Ages and the Renaissance, hermeneutics emerges as a crucial branch of Biblical studies. Later on, it comes to include the study of ancient and classic cultures.

<sup>13</sup> As presented by Bonjour (2002) and Hacking (2004), on the one hand, *epistemology* is the study of knowledge and justified belief; on the other hand, *ontology* is the philosophical study of the nature of being, existence, and reality, as well as the basic categories of being and their relations. The *epistemological tradition* mainly stretches some ways of dealing with issues such as the nature and scope of knowledge, the necessary and sufficient conditions for it, the sources and structure of it, and the limits of it. The *ontological tradition* deals with questions concerning what entities exist or can be said to exist, and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences.

communication is categorised as content preserving and the communicators do not challenge the ideas contained in the communicated messages by means of any inferential process. Romero-Trillo (2010) assumed that this approach marginalised the context that is created by the speaker and the listener in favour of what is simply believed to be true as what is understood. However, how to understand understanding and how to assess or evaluate understanding still remains a complex question that not only involves philosophical but also linguistic and psychological structures and reconstructions behind it.

From a cognitive perspective, understanding is a slippery conceptualisation, and speakers can only rely on the other's certification of having or not having understood a certain concept or proposition in order to attest to the understanding of an idea (Romero-Trillo, 2010). The cognitive approach studies the relationship between cognitive processes and meaning. However, Weigand (1999) pointed out that understanding cannot be viewed as a cognitive process, because the speaker is not aware of the misunderstanding at the time when it occurs. Understanding and misunderstanding must therefore be communicative or interactional rather than cognitive. There would be no way of telling the difference between understanding and misunderstanding by looking only at the cognitive process—they are the same. Thus, the unfolding discourse on an interactional level is needed to identify that a misunderstanding has occurred (see more discussions in the following section).

From a psychological perspective, Bruner (1990) has stated that the central concept in human psychology is *meaning* (p. 33) and the process and transaction involved in the construction of meaning is *understanding* (p. 33). This psychological approach is also cognitive. Because no one could possibly understand or access what others really understand deep down in their minds, *overt* (i.e., publically expressed) understanding is the focus of the studied phenomenon of understanding in general. The comprehension and learning of knowledge, for instance, in terms of connecting facts, relating the newly acquired knowledge to what is already known, and using it to solve both old and new problems, has become important when assessing understanding in psychological and pedagogical studies (e.g., Mintzes, Wandersee, & Novak, 2005; Fisher & Frey, 2007).

From a pedagogical perspective, in the domain of teaching and learning, Nicker-son (1985) claimed that to understand something does not only mean “having the knowledge of it” but also “doing something with it” (p. 234). Understanding is often accounted for as an outcome of, an objective in, or a prerequisite of learning (Säljö, 1982). Among other things, for example, how learners implement knowledge in practice (in particular in technical training), understanding has been frequently assessed through the learner's performance in a written test or oral questions (even in

technical training schools). It shows that in both pedagogy and psychology research, the assessment of understanding involves more or less linguistic devices in both spoken and written form.

This thesis aims to investigate certain linguistic devices and explore how understanding is conveyed through or in relation to them. As the linguist Ziff (1972, p. 20) suggested, “to understand understanding is a task to be attempted and not to be achieved today or even tomorrow”; the present study makes some attempts and contributions in order to understand more about understanding (in relation to some particular linguistic devices) in human social interaction in real-time.

## 2.4 Interactional approach to studying situated understanding in real-time

Real-time understanding can be also viewed from an interactional perspective. Conversation analysis researchers, for example, Sacks, Schegloff, and Jefferson (1974), view understanding as an interactive and situated phenomenon shown by the sequential use of language. In this perspective, understanding is something visible in discourse and is negotiated by the interlocutors through speech and *contextualisation* (a term first proposed by Gumperz, 1982) of *relevance* (see later the *relevance theory* of Sperber & Wilson, 1986).

*Context*, according to Linell and Thunqvist (2003, p. 410), originally refers to “what co-occurs with the text but is outside the text itself” and “what is necessary as a complement” for anyone “to make sense of the text”. Interlocutors are not isolated message senders and receivers, but information holders and sharers of everything surrounding the interaction such as the social interrelations, activity situations, and all the other “relevant contextual conditions” that “are assumed to be dynamically activated and accomplished” in interaction situations (see also Linell & Thunqvist, 2003, p. 410; Linell, 2018).

Generally speaking, contextualisation refers to the use of language and discourse to signal relevant aspects of an interactional or communicative context (Couper-Kuhlen, 2001). Contextualisation cues function by indexing or evoking interpretive schemas or frames within which inferential understanding can be achieved (Gumperz, 1982; Tannen, 1993). In other words, the contextualisation process simply means that the communication context affects the interlocutors’ understanding and expectations of each other’s communicative actions, and perhaps also the analyst’s interpretations of the communication itself and the interlocutors’ communicative actions.

This corresponds to Clark and Schaefer's *contribution theory* (1989), which proposes that in a dyadic conversation a speaker assumes that what (s)he has said is established as sufficiently understood only when the other has acknowledged it either by an explicit confirmation or by contributing a relevant subsequent utterance. Interpretation of the conversation has to be based on not only the explicitly expressed behaviours but also the conversation situation and context. Weizman (1999, p. 843) emphasised the notions of *shared* intentionality, which refers to the joint purpose of developing the interaction collaboration, and *collective action* that refers to the interactive behaviours involved in interaction and the interaction itself. Collaboration and interaction between the interlocutors (*we-level*), and within each interlocutor him/herself (*I-level*) are both important in terms of assigning a relevant meaning to the communicated information. Lynch (2011) further suggested that such an interactional approach to studying understanding is predominantly conversation analytic and informed by logical and grammatical investigations. He stressed that the meaning of the human's social action was not simply attached to the linguistic forms of the description of it, but implied and influenced by the conventions, interests, and knowledge of relevance.

*The relevance theory* was first proposed by Sperber and Wilson (1986), and seeks to explain the implicit inferences in understanding communication. It states that the hearer will search for meaning in any given communication situation, and once he or she finds a meaning (of relevance) that fits his or her expectation (of relevance) he or she will stop processing. Relevance is widely seen as a premise of understanding since understanding is a necessary requirement for a speaker to make a relevant contribution to the cooperative communicative interaction.

Allwood (1986) has stated that an initial characterisation of understanding is that it is a relation between "an agent that can understand and some particular information that is understandable and is being understood" (p. 2). He stressed that relevance is an indicator of understanding in interactions between utterances. Dascal and Berenstein (1987) have also emphasised that understanding an utterance means to understand the communicative purpose and the things the uttered affair is related to. To Dascal and Berenstein, understanding is always pragmatic understanding, which is somewhat similar to Allwood's idea, as mentioned above. Linell (2009) has further stated that understanding involves an interaction between the linguistically conventional meaning and the speaker's intended and exhibited meaning as well as other relevant contextual information and meaning. Although the predominant conception of understanding in philosophy and classic social theory is interpretive, the interpretation of understanding in particular in social interaction accounts more for the *relevance interpretation* than merely interpretation (see Lynch, 2011, pp. 551–

554). That is, understanding in communication has much to do with relevance in context, topic, and information sharing. When people have understood enough of what has just been going on of relevance, they can proceed to the next step in the communication or practical activity.

This interactional approach differed from Searle's *monologic speech act theory taxonomy* (1975, p. 345), in that it paid attention to action and reaction and all types of speech acts were regarded as dialogically oriented speech acts (i.e., either initiative or reactive). It was later developed as a *general dialogic principle* (Weigand, 1999, p. 35) and *dialogism* (Marková, 2003, p. 16; Linell, 2009, p. 19). Weigand (1999) seems to think that utterance acts are either initiatory or reactive; however, Linell (1998, 2009) holds that each utterance has both responsive (reactive) and projective (initiatory) properties although in different proportions.

The present study adopts the interactional approach in the sense that communicative understanding expressed in each utterance has both responsive and projective properties in relation to the earlier and the following utterances and their manifested understanding actions, respectively. Also, the specific (local) understanding action manifested in each utterance contributes to some kind of holistic (global) understanding sequentially in and through discourse sequences. Understanding is a dynamic interaction between interlocutors. It develops throughout the entire interaction activity. On the one hand, one may argue that there is basically one core understanding of how much the buyer should pay in a business negotiation conversation, one core understanding of what to make for dinner in a couple's telephone talk, and also one core understanding of where to spend the coming vacation in a family dialogue. On the other hand, all the interaction turns and contributions are full of materials for understanding. One has to choose where to set the line between holistic (global) understanding (of a long sequence of context) and specific (local) understanding (of a short sequence or even an utterance or a word). What constitutive components of understanding are there in communication requires research. This thesis work attempts to address this issue.

Also, as Allwood (1986) has said, when it comes to research understanding in language and communication, "we can view understanding both as a process and as a state" but "we have a preference linguistically for viewing it as a state", which is supported by "the factive nature of understanding" (p. 10). That is, understanding is linguistically preferred as a static relation to a fact that is often information presented in the earlier utterance. Although this thesis focuses on investigating the understanding actions and reactions of the interlocutor who is mainly playing a role as listener, the speaker's communicative intention and meaning and implicature as well as the

contextualisation of relevance and the interactivity<sup>14</sup> between speaker and listener (e.g., action and reaction) are all taken into account in the analysis.

## 2.5 Defining understanding: sense-making, meaning, and shared understanding

According to Samovar et al. (2012), people encounter an overwhelming amount of varying information that they must cognitively process and assign meanings to, and this process of selecting, organising, and evaluating the information is referred to as *perception* (see p. 13). *Understanding* (see Dodd, 2005, p. 65) in a broad sense primarily refers to the process of comprehending or interpreting the perceived information. As soon as people perceive some information, they start making sense of it and understand it with regard to some relevance and values (Sperber & Wilson, 1986; Zlatev, 2009). Being both part of the *sense-making* process, *perception* and *understanding* always occur interdependently. *Sense-making* includes all levels of awareness, consciousness, and intentionality, whereas *meaning-making* has more to do with what is consciously meant and intended (see Zlatev, 2009). Understanding can be seen as the process of making sense of the perceived information in association with (inference of) the meaning of the information, and it can be regarded as a form of sense-making.

Marková (2003) and Linell (2009) have made a strong claim that any action must be meant and understood, and meaning and understanding are absolutely central concepts when approaching the *dialogue*<sup>15</sup> and the world. Both meaning and understanding are achieved in situated interactions. “Language is not primarily a language of representation; rather, representing something can be reanalysed as a kind of action”, and *interaction* is basic to this action (Linell, 2009, p. 221).

---

<sup>14</sup> In this thesis, the level of *interactivity* (e.g., high or low) is measured by means of the number of utterances and contributions, including both actions and reactions of the discourse. If there is a great number of utterances and contributions in the interaction, the interactivity of this interaction is regarded as high.

<sup>15</sup> Marková’s (2003) and Linell’s (2009) *dialogical theories* use “*dialogue*” in a more abstract sense, as phenomena exhibiting dialogicality, which is taken to mean roughly other-orientation in sense-making. *Dialogue* (sociodialog) is not limited to dyadic or FTF interaction characterised by symmetric and cooperative communication among equals. In this thesis, the term *dialogue* is used in a more restricted way. Namely, it is an interdependent, sequential, and co-constructed interaction between two or more mutually co-present (not necessarily face-to-face) participants (see Luckmann, 1990), who share and coordinate factual, emotive, and volitional information for more or less joint purposes (see Allwood, 2000). This “concrete” form of dialogical processes (sociodialogue) is often spelled in the American way (*dialog*).

## 2.6 Overt understanding in social interaction

In social interaction, people want to communicate and exchange their beliefs, values, and thoughts for diverse reasons such as needs, interests, and wishes, and at the same time people want to find out whether their communicated information is perceived and understood, preferably, in a similar or same way as intended and anticipated. It is likely that, normally, everyone wants to be perceived and understood in an effective and hopefully pleasant and friendly way. Communication that serves the communicative purposes is preferable to *miscommunication*<sup>16</sup> (Mustajoki, 2012, p. 216). However, in reality, complete understanding and perfect communication are too ideal to aim for, and it is not uncommon that people are understood in many different ways and there are various types of understandings.

The concept of understanding has a key position in the study of social interaction. Allwood (1986) claimed that it was reasonable to regard social interaction or communication as, on the one hand, the production of and, on the other, the understanding of communicative behaviours. Understanding of social communication is just like knowledge about the world, neither of which is absolutely certain. The quest for *absolute certainty* (Linell, 2009, p. 227) of sense-making and understanding, often at the top of philosopher's agenda (e.g., Descartes), is unrealistic; instead, a *relative certainty* (Linell, 2009, p. 227) regarding how to continue for the current purposes of communicating (see Garfinkel, 1967; Bakhtin, 1986; Taylor, 1992) is more practical. Understanding a communicative activity, for example, understanding in social interaction, by using Linell's (2009) words, is like understanding a piece of visual art that involves three aspects, that is, what in the outer world the picture depicts, what the role of the artist is in creating the picture, and what the role of the viewer is in seeing it from a certain aspect. The truth is that people simply do not disclose everything that they have in mind and some cognitive processes cannot be brought into language in a *completely accountable* (Linell, 2009, p. 15) manner. Therefore, what language and communication researchers can observe and investigate and then make interpretations of is restricted to what is manifested or exhibited through language

---

<sup>16</sup> According to Mustajoki (2012), *miscommunication* is defined as the situation in which the recipient understands the message in a different way than intended by the speaker. In this thesis, *miscommunication* is used to refer to unsuccessful communication, in which insufficient understanding or an understanding problem occurs and sufficient understanding is not achieved as intended and the communication goal is not accomplished (see definitions of *(in)sufficient understanding* and *understanding problem* in Section 2.8).

communication (i.e., *overt understanding*), although in fact immediate understanding is often quite *shallow* and sometimes the listeners may *fake understanding* (Linell, 2009, p. 271).

Because people cannot have direct access to other people's minds or cognitive processes to see the state or process of their understanding of the communicated information, the evaluation of understanding is usually based on how understanding is made manifest through the relevant communicative behaviours. It is a methodologically difficult issue to identify whether the interlocutor B really understands what the interlocutor A means, or B is just showing that B understands something, or B (and A) think(s) that B understands. As discursive psychology and conversation analysis reasoned methodologically, the *overt interaction* (Linell, 2009, p. 20) is the best and often the only place where cognitive processes such as understanding can be observed. Thus, *private understanding* (Linell, 2009, p. 230) that is not accessible to the analyst and possibly not to the other interlocutors either is not taken into account in this thesis. Instead, the focus is on understanding, which is overt and manifested through communicative behaviours in interaction (see Clark & Schaefer, 1989; Macbeth, 1994; Lindwall, 2008).

As both *gestural*<sup>17</sup> aspects of communication and the prosodic aspects of speech add value to the *vocal-verbal* communication (Goodwin, 2018) and they supplement or modify each other in the process of meaning and understanding sharing in communication (Mitchell & Ross, 2013), the auditory vocal-verbal cues including both the speech and its prosody as well as the visual gestural ones are considered to be equally important in social interactions. Thus, in this thesis, understanding and its signal in social interaction will be evaluated from vocal-verbal, gestural, and prosodic aspects.

## 2.7 Partially shared understanding as a basic idea

Understanding is philosophically a partially shared understanding which has been mentioned by a large number of researchers (Garfinkel, 1967; Rommetveit, 1974; Bakhtin, 1986; Taylor, 1992; Linell, 2009). According to them, from a philosophical perspective, understanding is an incomplete and unfinalisable dynamic process.

Rommetveit (1974) has regarded interaction as the “temporarily shared social reality”, which is created and developed among parties in a communication situation (see Linell, 2009, p. 97). Understanding is closely linked to responding to situations and utterances, and we do not need to bother about *understanding everything* (Linell,

---

<sup>17</sup> In this thesis, communication has both *gestural* and *vocal-verbal* aspects. More definitions and discussions of them and the related are presented in Section 2.11.

2009, p. 222). Rommetveit (1974) has defined situated understanding as dialogically constituted and only partially shared. Bakhtin (1986) has pointed out that understanding is part of an *unfinalisable*<sup>18</sup> sense-making (p. 63), if we look upon it beyond the boundaries of the particular situation. Garfinkel (1967) and Taylor (1992) have stated that we need understandings only for “*current practical purposes*<sup>19</sup>”, that is, a “partially shared and shallow understanding” but sufficient for us to go on with our current doings (see Linell, 2009, p. 222). Understanding one another in a real communication situation is not a matter of achieving *complete and completely shared understanding* but typically of achieving some *partial or shallow understanding* for the practical purpose of being able to continue with what is currently going on (Linell, 2009, pp. 226–227). That is, almost all types of understanding in communication are partially shared, where sufficient understanding serves the practical communication purpose. Since there is almost no complete understanding in this world and the overwhelmingly majority is *partially shared understanding*<sup>20</sup>, only this “partially shared understanding” is the “understanding” that researchers in language and communication are investigating and every ordinary individual person is coping with in social life and communication.

## 2.8 Operationalising and classifying understanding

According to Lynch (2011), understanding has been treated as a methodological prerequisite of any analysis of human action. Although understanding is fundamental in human communication and social activities, there is hardly any classification of understanding that is simple to comprehend and easy to utilise. Many researchers

---

<sup>18</sup> Bakhtin (1986, p. 63, p. 66, p. 170) has proposed the theory of *unfinalisability of dialogue*, suggesting that dialogic expression is *unfinalisable* and always *incomplete* and *productive* of further chains of *responses*: *meaning* is never closed and always oriented toward the future. The unfinalisability of dialogue cannot be overcome in thought or speech. Individual people cannot be finalised, completely understood, known, or labeled.

<sup>19</sup> *The practical purpose* generally refers to sharing and making sense of the information presented; it can be carrying on a conversation as primarily addressed by Garfinkel (1967) or carrying on a specific communication and learning task as presented in Lindwall and Lymer (2011) and also in this thesis.

<sup>20</sup> The term *partial understanding*, which is used later in Study 2 in this thesis, refers to a particular type of understanding. It is as a different term separate from any philosophical concerns like the above-mentioned. It has been used by a few researchers, for example, Allwood (1986), to refer to some type of understanding problem or difficulty. Because of a similar conceptual purpose and meaning, the term *partial understanding* is also used in this thesis to refer to a type of insufficient understanding. Details of its definition and conceptualisation will be presented later.

have concluded that analysing and categorising understanding has become a methodological procedure:

understanding might turn out to be what the prior speaker intended, or it may not; whichever is the case, that is something that gets displayed in the next turn in the sequence. We described this as the *next-turn proof procedure*, and it is the most basic tool used in CA to ensure that analyses explicate the orderly properties of talk as oriented to accomplishments of participants, rather than being based merely on the assumptions of the analyst. (Hutchby & Wooffitt, 2008, p. 13)

Obviously, the term “turn” can be interpreted as either a specific *turn management* or *turn-taking* or a further followed-up communication context in general. It makes sense to adopt the latter, because *understanding* is not only or necessarily proved in the direct “next turn” in the conversation. For instance, it is not uncommon that it can take a while, for example, a few utterances or “turns” or even a lengthy talk, for the interlocutors to be able to understand the earlier communicated information or realise if anything has been misunderstood. Again, understanding should be studied from an interactional and contextualised perspective.

### 2.8.1 Challenge and opportunity: analytical interpretation

The operationalisation of understanding is, however, problematic, in that interpreting or annotating the manifestation of understanding in relation to the communication context is usually indirect and somehow subjective to the annotators and researchers.

We must, in fact, to a great extent, rely on interpretation of the content and function of linguistic data in order to find cases which might throw light on the different states and processes of understanding. This, in turn, means that the sense in which we are using ‘operationalization’ and ‘observation’ are far removed from the ways in which these terms were used in the classical days of logical empiricism. (Allwood, 1986, p. 23)

The most common approach is to use introspection or observation of the ways in which this abstract phenomenon of understanding is manifested through communicative behaviours, by using the evaluation criteria in terms of “relevance, coherence, conventional correctness, correspondence, consensus, degree of completeness, and degree of activation” (Allwood, 1986, p. 24). As a consequence, there is no escaping that a fairly high degree of interpretation is present in the empirical observational data (Allwood, 1986). This again stresses that it is a methodologically difficult issue

to identify whether the information is understood or not. Thus, understanding is interpreted and studied from an analyst's (analytical) perspective in this thesis, though tests of inter-rater reliability were carried out to make this less subjective.

Furthermore, it is even more difficult to distinguish the different levels of how much information is understood or not in the real communication practice. Wittgenstein (1958, p. 61) proposed that we should try not to think of understanding as a 'mental process', for that is the expression that confuses us; instead, we should ask ourselves in what sort of case and circumstance we say 'now I know how to go on'. In Clark and Schaefer's *Contributing to Discourse* (1989) and Clark and Wilkes-Gibbs' collaborative process (1986), they have also pointed out that interlocutors try to establish a mutual belief for each utterance that the addressee has understood what the speaker meant well enough for *current purposes* (Clark & Schaefer, 1989, p. 259) for developing more common ground for further discourse. Clark, Schaefer, and Wilkes-Gibbs all claimed that during the discourse process, there were different levels and stages of understanding prior to an achieved and exhibited shared understanding. However, Clark, Schaefer, and Wilkes-Gibbs did not make any effort to provide a concrete criterion of what the levels and stages of understanding are and how to identify and measure them. Firstly, this is probably because theoretically it is not easy to create such a measurement or scale; for instance, understanding can be understanding nothing, a little, some, plenty, a lot, most, all, and so on. As Verdonik (2010) said, people can never understand even the same discourse in an identical way. It is hardly possible to make a universal detailed measurement. Secondly, it is probably because even in practice there is a scale of such complexity that it is still very difficult to use it to annotate the empirical data of interactions and achieve sufficiently high reliability. Nevertheless, a number of researchers have agreed that communication analysis in general has an important part to play in developing the understanding of understanding in interaction, and that an empirical analysis of the process of understanding in interaction can make a significant contribution (cf. Anderson, 2006).

## 2.8.2 Influential ontological criteria for classifying understanding

Allwood (1986) proposed ontological criteria for evaluating different types or degrees of understanding, based on the notions similar to that of the degree of understanding such as *pseudo concept* (Vygotsky, 1962, pp. 66–69), *depth of intention* (Naess, 1966, pp. 33–36), and *depth of understanding* (Allwood, 1976, pp. 71–73). Allwood's (1986, pp. 11–20) criteria include *total lack of understanding*, *awareness but no understanding*, *partial understanding or lack of understanding*, *insufficient un-*

*derstanding, incorrect understanding, incomplete understanding, and misunderstanding.* Obviously, from a semantic perspective, it is not easy to include partial understanding, insufficient understanding, and incomplete understanding at the same time within the same taxonomy. Also, as Allwood (1986) admitted, it is not easy to separate *incomplete understanding* and *insufficient understanding* from *partial understanding* and it is difficult to separate *incorrect understanding* from *misunderstanding* (see p. 20). In other words, it is not easy to set up a criterion to evaluate the degrees of understanding on a micro level although it may be possible in a more general sense. Therefore, a simplified criterion for evaluating understanding, which includes *sufficient understanding, misunderstanding, non-understanding, and partial understanding*<sup>21</sup>, is used in the present thesis. These types of understandings will be discussed below.

Weigand (1999) suggested a categorisation of understanding that includes (harmonious) understanding, misunderstanding, and non-understanding as follows.

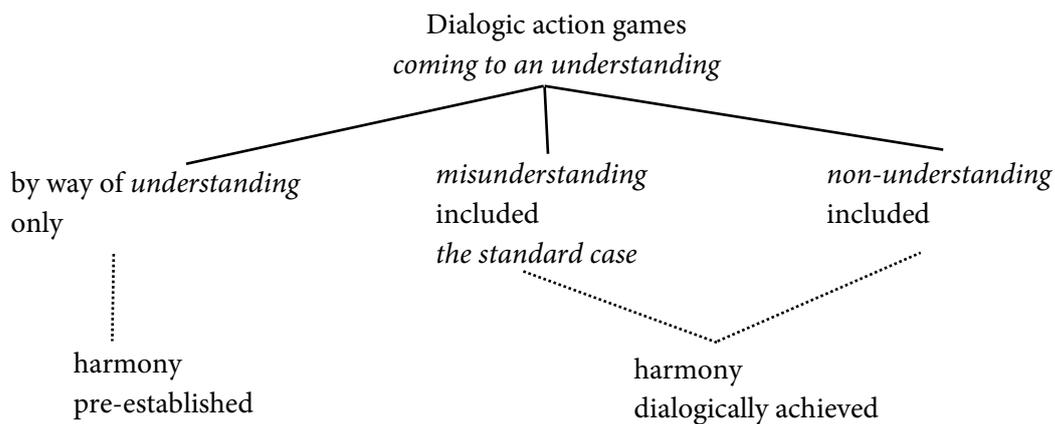


Figure 2.1. Dialogic action games: coming to an understanding (cited from Weigand, 1999, p. 771).

Weigand's (1999, pp. 770–771) taxonomy of dialogic action games (see Figure 2.1) explicitly treats the issue of understanding as a process of *coming to an understanding* rather than being a static understanding. In her theory, misunderstanding usually occurs without the interlocutors' awareness, and it is normally corrected in the on-

<sup>21</sup> The first three categories of understanding are used in Study 1, while all four categories of understanding are used in Study 2.

going discourse. In contrast to misunderstanding, the hearer identifies non-understanding. The hearer signals the understanding problem and thus initiates the process of clarification, and the speaker makes the clarification. However, Weigand's model may be restricted in some way. For instance, *understanding (harmony pre-established)* is not necessarily to be *pre-established*; instead, it is usually *dialogically achieved* and developed. *Misunderstanding* may be or may not be corrected. The latter may actually lead to more misunderstandings in the further discourse. *Non-understanding* may be regarded as one stage of understanding, which might be developed into sufficient understanding but not necessarily always. Sufficient understanding, misunderstanding, and non-understanding are different types of understandings that are distinct from one another. Thus, do *misunderstanding* and *non-understanding* necessarily lead to *coming to an understanding*? If so, why and how are they constituted by the process of *coming to an understanding*? Can *misunderstanding* always or normally be corrected and then build up further *understanding*? Maybe some misunderstandings are fruitful and constructive, but how to distinguish them from those that are not? There may be some interaction existing between different types of understandings, but what is it? Perhaps, there are no absolute answers to these questions, due to the fact that the world is one of high complexity in the relationship between what is meant and what should be understood and also that there are always differences in cultural, linguistic, physical, and cognitive abilities from one interlocutor to another. For example, when analysts want to investigate the cognitive and emotional mechanisms of human actions, they are always confronted by a risk of "misunderstanding" and "non-understanding". As Weigand (1999, p. 781) herself claimed, "we cannot start from a model of understanding an utterance as an ideal construction in which all points are fixed (as in Chomsky's ideal grammar) or in which understanding is presupposed (as in a certain type of dialogue grammar)", we have to view understanding from an interactional and contextualised perspective. Only the communication context and the relevant behaviours can exhibit what is understood or not and how it is understood. Preferably, all the aspects of communicative interactions (e.g., vocal-verbal, gestural, and prosodic features) are taken into account.

Based on Allwood's (1986) and Weigand's (1999) suggestions for classifying understanding and Linell's (2009) suggestion of partially shared but sufficient understanding as well as the fact that understanding is closely related to *response* and *anticipation* (see also Bakhtin, 1986, p. 69), the criterion used in this thesis is based on

whether the information is understood sufficiently well and correctly<sup>22</sup> in relation to what is required to continue the conversation, sometimes including what can be inferred about the speaker's intention and anticipation.

### 2.8.3 Categories of understanding

In this thesis, the following categories are used: *sufficient understanding*, *misunderstanding*, and *non-understanding*, which are further described below.

#### Sufficient understanding

*Sufficient understanding* refers to the partially shared understanding that is sufficient to serve the current practical purposes (Garfinkel, 1967; Bakhtin, 1986; Taylor, 1992; see Linell, 2009, p. 222) of information sharing, sense-making, and continuing communication, no matter how much is partially shared (see Linell, 2009, p. 39, p. 222, and p. 227). Also, the continuation of the interaction seems *smooth*<sup>23</sup> without a manifested *speech act* of questioning, although there can be doubts, arguments, and disagreements. Sufficient understanding is usually exhibited in the speech acts of declaring and persuading. In addition, sufficient understanding in the present study is also defined as the information being understood in a way that is correct for current purposes in relation to what the relevant communicator intends and anticipates, no matter how much is correct. That is, sufficient understanding occurs when the interlocutors are content with the understanding of one another and it is good enough to proceed further in the communication (see also Lindwall & Lymer, 2011). Theoretically, there can be cases in which the participant's understanding is more sophisticated or better than the speaker's, they are also defined as sufficient understanding.

#### Misunderstanding

*Misunderstanding* is defined as a form of insufficient understanding in this thesis. Misunderstanding only occurs when the information is understood in an incorrect way or deviates from what is intended or anticipated (see similar use in Weigand,

---

<sup>22</sup> *Sufficient* means that it is enough for the current practical purposes of understanding what has just been going on and knowing how to proceed in the interaction and keep the interaction going on and on from moment to moment. *Correct* means that the information is understood in the same or a similar way as is intended and anticipated by the speaker, which is good enough for carrying out the current purposes of interaction.

<sup>23</sup> In this thesis, the quality of a *smooth interaction* is entirely determined by the correlation of interests between the interlocutors (see also Lu, 2017); the interlocutors perceive and accept or reject the communicated information without putting forward questions or challenges.

1999) although it can perhaps serve the current communication purposes of exchanging information and carrying on the conversation. Misunderstanding may occur as other insufficient understanding problems (see later) during the process of achieving sufficient understanding, although it does not always achieve sufficient understanding. Instead, it can sometimes cause further misunderstandings and may not lead to anywhere close to sufficient understanding. Identification of misunderstanding is more contextual relevance dependent than other forms of understanding problems (Lu, 2014); thus, it is more observable in task-oriented activities than in everyday conversations.

In a case of misunderstanding, the interlocutor may think he or she understands the communicated information, but in and through the ensuing utterances and behaviours it is found that he or she in fact has not understood or not aligned with the other interlocutor's intentions. That is, the interlocutor who misunderstands is not aware of it (see also Weigand, 1999). In fact, misunderstanding is often not *detected or even detectable*<sup>24</sup> by the participating interlocutors. The interlocutors just carry on the interaction without being aware of the occurrence of misunderstanding throughout the entire interaction. This is a very interesting empirical phenomenon, yet it is not easy for the analyst to explore the deeper reasons and provide explanations as to how the interlocutors could carry on the interaction without becoming aware of such problems of misunderstanding.

## Non-understanding

*Non-understanding* is another form of insufficient understanding. Based on Weigand (1999), non-understanding in this thesis is identified when the information

---

<sup>24</sup> *Detect* (Schegloff, 1992, p. 1320; Bazzanella & Damiano, 1999, p. 823; Dascal, 1999, p. 754; Weigand, 1999, p. 765; Verdonik, 2010, p. 1377) and its synonyms *notice* (Schegloff, 1992, p. 1299; Hirst, McRoy, Heeman, Edmonds, & Horton, 1994, p. 215; Bazzanella & Damiano, 1999, p. 821; Weigand, 1999, p. 768; Verdonik, 2010, p. 1367; Mustajoki, 2012, p. 234), *recognise* (Schegloff, 1992, p. 1322; Hirst et al., 1994, p. 215; Bazzanella & Damiano, 1999, p. 823; Dascal, 2003, p. 295; Dascal, 1999, p. 754; Weigand, 1999, p. 776; Verdonik, 2010, p. 1377; Mustajoki, 2012, p. 228), *identify* (Bazzanella & Damiano, 1999, p. 822; Dascal, 1999, p. 758; Weigand, 1999, p. 770), and *discover* (Hartnack, 2013, p. 65; Dannerer, 2004, p. 105) have been widely used in the studies of language communication and understanding. The term *detect* is adopted and employed in this thesis, because from my point of view, compared to other terms, *detect* has more to do with active attention and participation of the interlocutors (see more reasoning in Posner, Snyder, & Davidson (1980)).

In this thesis, *detect* and its derivatives are used from a participant's perspective because understanding problems are almost always analytically detectable and detected. What this thesis focuses on is describing from a participant's perspective what has actually happened between the participants during the interaction and investigating the participants' relevant communicative behaviours from an analytical perspective.

is not understood at all for reasons such as lack of access to the information itself or some background knowledge of relevance. As the opposite of sufficient understanding, non-understanding cannot serve the current communication purposes of sharing and making sense of the information presented. Normally, non-understanding is manifested in the speech act of questioning, thus it is not difficult for the interlocutors to detect and resolve non-understanding problems; sufficiently shared understanding can usually be achieved eventually.

## 2.9 Handling of understanding problems

Earlier theorists have identified several ways in which people handle and solve understanding problems in communication. Concepts such as *refashion*, *meaning repair*, *common ground*, *negotiation*, *plan derivation*, *elaboration*, *reference repair*, and *interactive alignment* have been developed and used by different researchers (as follows).

### 2.9.1 Meaning repair

Clark and Wilkes-Gibbs (1986) have proposed that after one participant presents an initial referring expression the other participant would make a judgment on accepting or rejecting it or postponing the decision about it. If the referring expression is rejected or the decision is postponed, one of the participants would initiate *refashioning*, the referring expression for *meaning repair* (see Clark & Wilkes-Gibbs, 1986, pp. 6–7). This takes place in the form of either adding further qualifications of the referring expression or changing the original referring expression to a new referring expression. The referring expression will be judged again, and the same process continues until a referring expression is accepted for the current communication purposes. This final referring expression becomes part of the participants' *common ground*.

Often, interlocutors are engaged in a kind of *negotiation* (Clark, 1993, p. 3; Clark & Wilkes-Gibbs, 1986, pp. 2–3; Horton & Gerrig, 2005, p. 3) in order for one of them to understand a reference that the other wishes to make. This negotiation can be a factual negotiation, which primarily concerns the domain of the reference, or a meta-interactive negotiation that mainly concerns the interlocutors' rights and obligations (see e.g., Moeschler, 1985; Roulet, 1992).

Referring expressions are represented by *plan derivations* and *elaborations*, and an unsuccessful referring expression is an invalid plan for *reference repair* in the participants' collaboration (Hirst et al., 1994, p. 216, p. 220, p. 223). Pollack (1990) has claimed that all the plan derivations will be evaluated and judged by the participant,

and when a valid one comes up the participant will believe he or she has identified the reference and has understood it. Thus, the common ground of the participants is updated with the accepted referring expression. Otherwise, the possible constraints and sources of the *invalidity* (Pollack, 1990, pp. 79–80) will be noted by the participant and used in further attempts to fix the understanding problem. Hirst et al. (1994) have put it differently:

if a participant does not notice anything unusual, she may assume that the conversation is proceeding smoothly. But if she hears something that seems inconsistent with her expectations, she may hypothesize that there has been a misunderstanding, either by herself or the other, and produce a repair- an utterance that attempts to correct the problem. (Hirst et al., 1994, p. 223)

Pickering and Garrod (2006) characterised this process as *interactive alignment* (p. 203) of meaning and understanding and regarded it as the basis of successful communication. A combination of personal and social accounts of discourse can help to identify, acknowledge, and repair understanding problems and thus support negotiation of meaning.

In this thesis, the term *meaning repair* is used, referring to a communicative action that attempts to correct the understanding problem that has occurred and been observed during the process of *meaning negotiation* in order to achieve anticipated information sharing. Meaning repair is an important part of sense-making and information sharing.

## 2.9.2 Initiation of meaning repair: self- or other-initiated

Hirst et al. (1994, p. 216) have presented a model of *initiator* and *recipient* of the collaboration of handling non-understanding problems. It concerns the question of who initiates and who receives the meaning repair, which is primarily a question of the initiation of meaning repair.

This understanding negotiation and meaning repair can be *self-initiated* or *other-initiated* (cf. Schegloff, Jefferson, & Sacks, 1977; also, see *self-misunderstanding* and *other-misunderstanding* in Hirst et al., 1994, p. 226).

In this thesis, self-initiated meaning repair refers to the repair, initiated by the speaking person, of the information that causes understanding problems or difficulties. In this case, the speaking person is the one that detects the understanding problem in the first place. Then, this speaking person initiates meaning repair and negotiates for an intended and anticipated shared understanding. Self-initiated repair is usually in the form of a statement, often with a purpose of modifying or adding more

sufficient and necessary information in order to promote sense-making and information sharing.

By contrast, other-initiated meaning repair refers to the repair initiated by the listening person, who has not sufficiently understood the information communicated and who wants further clarification. Other-initiated meaning repair is usually in the form of questions and its purpose is still to achieve a sufficiently shared understanding.

### 2.9.3 Performance of meaning repair: self- or other-performed

After being initiated, meaning repair is performed sequentially in the discourse. In fact, the person who primarily performs the meaning repair or says who plays the dominant role in the repair and correction work may not be the one who has initiated it. According to Schegloff, Jefferson, and Sacks (1977), meaning repair can be self- or other-performed, namely self-repair or other-repair, depending on who does the actual repair work. This actual repair work can be done by either of the participants. Schegloff et al. (1977) have claimed that self-repair dominates other-repair in discourse and conversation, other-repair occurs frequently in adult-child interactions. In this thesis, meaning repair of understanding problems will be studied in terms of both its initiation and its performance, respectively.

## 2.10 Micro-feedback and understanding

As Bakhtin (1984) has said, we all live in a world of others' words. Everything one says or does is a response to someone or something (see also Lévinas, 1961). Response, responsivity, and responsibility, which are closely related in form and meaning, are what makes us human (Linell, 2009). When it comes to social interaction and communication, Schegloff (2007) has said that an utterance is never “said on its own behalf ... but on behalf of something else” (p. 260), which Linell (2009) called a *communicative project* (pp. 186–187) that is initiated by some participant(s) and sustained at least for a moment. No cognitive or communicative act (e.g., understanding) was randomly related to the environment, and every act (e.g., understanding) was selectively responsive to (a complex array of) contextual conditions that often include the particular communicative actions of others (see Schegloff, 2007; Linell, 2009).

### 2.10.1 Understanding: responsive to the prior and initiatory to the subsequent

Bakhtin (1986, pp. 69–91) has stated that any understanding of live speech is *inherently responsive*, the speaker is oriented precisely towards an *actively responsive* understanding, and any understanding is imbued with response and necessarily elicits further response in one form or another. He has stressed the responsive and interactive relation between words in conversation. That is, a word that responds also provokes responses with certain anticipations:

The word in living conversation is directly, blatantly, oriented toward a future answer-word: it provokes an answer, anticipates it and structures itself in the answer's direction... every word is directed toward an answer and cannot escape the profound influence of the answering word that it anticipates. (Bakhtin, 1992, pp. 279–280)

The listener thus becomes the speaker. According to Bakhtin (1986), the speaker talks with an expectation, for instance, of a response, agreement, sympathy, objection, execution, and so forth, and that each utterance is filled with various kinds of responsive and initiatory actions to other utterances of the given sphere of speech communication.

Clark and Schaefer (1989), similarly, stated that a *contribution to discourse* includes an initiatory presentation phase in which A is active, a responsive acceptance phase in which B presents understanding and (perhaps) A gives feedback to this later on, and so forth. This is regarded as a more *dialogical approach* (Marková, 2003; Linell, 2009) in the study of social communication. Clark (1996) proposed an underlying structure of action initiatives and responses that are regarded as the basic “interact” relations in the account of interaction. There is interdependency<sup>25</sup> between understanding and responding in that understanding is both a response to the prior contributions and a new initiative for the further contributions (see Goodwin, 1981; Schegloff, 1996).

This interactive feature of language use and understanding has been acknowledged by Linell (2009) in his depiction of the process of sense-making and understanding:

---

<sup>25</sup> In this thesis the level of *interdependency* (e.g., high or low) is primarily measured by means of the discourse responses. If there is a great number of responses in the interaction, the interdependency of this interaction is regarded as high. In the present study, interdependency is examined with a focus on reactions, whereas interactivity is measured with focuses on both actions and reactions. These two notions are interrelated in interaction analysis.

If speaker A utters something and thereby indicates a targeted understanding, then speaker B must indicate his understanding of this by some responsive action, typically another utterance, and then A has to show her reaction to B's response by yet another action (utterance). Note that without this third step, while A has access to B's understanding as displayed, or at least as partially displayed in his utterance, B has not yet received any reaction from A and hence cannot know whether his utterance, and his presupposed understanding of A's first utterance, fits with A's ideas; hence, no mutual and shared knowledge has been established (unless, of course, there are contextually established routines which make such checking procedures unnecessary). (Linell, 2009, p. 183)

As Linell (2009) explained, response shows the participant's understanding and stance with respect to the prior contributions in a dialogue; thus, understanding has a *responsive relation* (p. 357) to the prior actions in the interaction. Also, the responsive understanding tends to generate more dialogues, more interactions, and more contributions to discourse; accordingly, it has an *initiatory relation* (p. 179) to the possible subsequent actions in the interaction as well.

### 2.10.2 Micro-feedback: one typical responsive and initiatory behaviour in social interaction

Concerning how understanding is manifested through the responsive behaviours in interaction, Lindwall (2008) suggested one could look anywhere in the interaction for the evidence of understanding that is massively present (p. 217) throughout the interaction. Possibly, each utterance in the interaction shows an understanding of the prior utterance(s).

Clark and Schaefer (1989) identified different types of evidence of understanding that are graded roughly from weakest to strongest (p. 267):

- (1) *Continued attention*. B shows that he is continuing to attend and therefore remains satisfied with A's presentation.
- (2) *Initiation of the relevant next contribution*. B starts in on the next contribution that would be relevant at a level as high as the current one.
- (3) *Acknowledgement*. B nods or says "uh huh", "yeah", or the like.
- (4) *Demonstration*. B demonstrates all or part of what he has understood A to mean.
- (5) *Display*. B displays verbatim all or part of A's presentation.

Linell (2011) has addressed feedback and its functions (including understanding) as follows. A feedback item is by definition a recipient's response to a prior speaker's

contribution (utterance). A feedback utterance can be quite minimal or expanded, and the different types form a scale from very weak to quite strong ones. If a second-positioned speaker's contribution develops into a long turn, it would hardly be only feedback. Linell (2011, pp. 270–273) has proposed a scale of (mostly minimal) feedback items (from weak to strong) more or less like this:

- (1) Non-verbal only: noddings, gaze, facial expressions, manual gestures, bodily orientations, etc.
- (2) Minimal vocal-verbal items (normally accompanied by and integrated with “non-verbal” signals): *mm, uh-huh*, etc.
- (3) Single response particles: *yes, no, oh*, OK, etc.
- (4) Weak acknowledgement tokens (perhaps involving provisional agreements) (“I hear and I (provisionally) understand”): *I see, I understand, sure, yes certainly*, etc.
- (5) Stronger acknowledgements: assessments, strong agreement (presuppose but do not demonstrate understanding): emphatic *certainly, absolutely, good, too bad, brilliant, fantastic*, etc.
- (6) Repetition of (parts of) the prior speaker's utterance, *yes but ....* (own interpretation and formulation), *yes because ...* (demonstrate understanding, agreement extended)
- (7) (Partial) disagreement: *well..., no because ...*
- (8) Stronger disagreements (from the prior speaker's perspective, these can demonstrate *no* or partial and insufficient understanding)

This scale is a dynamic rather than a fixed system. For example, prosodies expressing more emotionality and engagement may move items from being lower-level to becoming stronger. Furthermore, the items mentioned above prioritise agreement with and understanding of the prior speaker. Disagreement and deviant understanding will necessitate stronger expressions, (i.e., resources on levels 7–8). Repair is sometimes linked to hitches in the interaction. According to Linell (2011), the dominant theories of conversation analysis (CA) argue that the weaker forms (categories 1–4) are just “continuers” or “go-ahead signals” (Schegloff, 1982, p. 71). They are normally shaped in the form of rather less outward-directed signals, not used by their speaker to try to take over the turn. Likewise, categories 1–4 will often be perceived as backchannel items (listener support items), rather than full contributions to the focused interaction (discourse). Therefore, it is usually assumed that the lower categories do not have the function of expressing (substantial) understanding of the prior

speaker's utterance. This is perhaps in line with what CA says about “claimed” vs. “demonstrated” understanding, the former likely expressed at levels 1–5 and the latter typically expressed at levels 6–8. However, a claimed understanding or a shallow understanding is nevertheless a kind of understanding that responds to the perceived message and projects the upcoming message and provides information to the other interlocutor about how to proceed with the conversation. For instance, if (s)he should elaborate the presented message in another way and make some meaning repair and correction, or if (s)he can leave the current topic with a good enough “shared understanding” and thereafter carry on the interaction and move on to the next topic.

### 2.10.3 Claimed understanding: often enough for all practical purposes in ordinary conversation

Because one cannot investigate all the aspects of how understanding is exhibited through all kinds of possible responsive or feedback behaviours in practice, Lindwall (2008) exclusively focused on the explicit utterances such as *I get it, I understand that, I see, yeah okay now I can go on*, and the like which signal understanding in lab work-based (science) education. These explicit expressions, as discussed also by Lindwall and Lymer (2011), “can be treated as pragmatic equivalents for expressing the Wittgensteinian sense of now I (or we) can go on” (see Lynch, 2011, p. 554).

Also, people in a discourse exchange can show that they understand and they can go on communicating without saying these explicit words but employing other linguistic devices. For instance, among other results, Lindwall (2008) has found that in ordinary conversation, it is often enough for all practical purposes to provide *minimal responses* (p. 249), such as *uhm* or *yeah*, to signal that one has understood what the other person has said. *Uh huh, mhm*, and laughter in Lindwall’s study, which are the same as or very close to the concept of *micro-feedback* in this thesis, are not only regarded as signalling “I’m listening” and “please continue” but also as evidence of specifically topic related understandings at a certain point of time in the interaction (see also Lindwall, 2008). How these minimal responses and understanding are related, for example, what specific micro-feedback is used in relation to which specific type of understanding, is not stated in Lindwall’s study (2008) but is investigated in the present thesis.

Just as importantly, Lindwall (2008) realised that even when a student simultaneously says *yeah, I get it* in the lab work, he or she may still not understand *what we’re supposed to do* (see p. 251) and “the lack of laughter is thus not primarily taken as a failure to understand..., but as a display of a potential need of further elabora-

tions ...” (p. 250). This simply suggests that not only the vocal-verbal expression itself but also the accompanying gestures and other associated features embedded in the communication context have a great deal to do with helping people to understand the understanding in social interaction. How can people possibly recognise whether the communicated information is understood, misunderstood, or not understood at all through micro-feedback? What kind of specific features does the micro-feedback expression have when it signals or correlates to different types of understandings? These questions highlight the importance of micro-feedback in social interaction.

## 2.11 Defining and operationalising micro-feedback

When people attend to or interact with other interlocutors, it is very likely that some information is perceived and reacted to; but it is possible that this happens without understanding or with only shallow understanding. Sometimes, the interlocutor takes an evaluative stance, such as agreeing or disagreeing. If the interlocutor agrees or disagrees, he or she must have understood what he or she agrees or disagrees about. However, the interlocutor can also deceptively and deliberately pretend to attend, perceive, and understand (such as *fake understanding* in Linell, 2009, p. 271). Fake agreeing or disagreeing are found infrequently in the discourse data, and is disregarded in the analysis.

As discussed earlier, although in real empirical data it is not easy to identify whether the interlocutor really understands or just shows or pretends that he or she understands, the overt and manifested understandings from an analyst’s perspective is the main focus and approach when studying understanding in real-time situated interactions. In this case, the coherent communication context and the details of the interlocutor’s responsive behaviours, for example, micro-feedback in terms of features of both modality (e.g., auditory and visual) and prosody (e.g., pitch and duration) which serve the contemporary communication purpose of keeping the communication going on, are the prominent helpers for identifying understanding.

### 2.11.1 Terminology

The phenomenon of micro-feedback has been studied in a number of terms, which include “listener response”, “continuers”, “go-ahead signals”, “signals of continued attention”, “accompaniment signal”, “concurrent feedback”, “acknowledgement token”, “response token”, “reactive token”, and “limited feedback” (cf. Deng, 2009). Terms like “minimal feedback” (Fishman, 1978), “backchannel”<sup>26</sup> (Yngve, 1970),

---

<sup>26</sup> The term *backchannel* occurs in earlier studies also in the form of *back channel* and *back-channel*.

and “feedback” (Nivre et al., 1992) have been used fairly commonly and widely in linguistic and communication studies.

Fishman’s *minimal feedback* refers to the minimal responses in short words such as *yeah, umm, huh*, and only those that excludes gestures (e.g. facial expressions and head movements) and expressions consisting of long phrases or even sentences, for example, *yeah I can believe it, okay you’ve started writing on your dissertation then?*, and *wow, really? It is fantastic*.

Yngve’s *backchannel*, according to Allwood (1993, p. 198), seemed to be characterised as “feedback giving out of turn” while “feedback giving in turn” and “feedback elicitation” were excluded. For Allwood, *feedback* includes all the “feedback giving and eliciting” means, both in and out of turn.

Nivre, Allwood, and Ahlsén’s *feedback* seemed to be used in broad sense of referring to communication responses as full contributions to the discourse (Wood, 2011, p. 85), for example, *how are you? I’m fine thank you...*, or comments and evaluations that are mainly made up of comprehensive and expanded responsive expressions (Mahboob & Knight, 2010, p. 31), for example, *well done you’ve made great progress but ...*, or some particular responsive linguistic device, for example, *yeah yeah yeah* or head nod that signals “I hear and understand what you have just said” or laughter that signals being amused or happy. In order to reduce or avoid this terminological ambiguity of *feedback*, the term *micro-feedback* is preferred.

The purpose of having this term *micro-feedback* is to highlight the pragmatic feature of being small in relation to understanding (e.g., in ordinary social interaction the relation is sometimes insubstantial or shallow) and the unobtrusive aspects of it in its semantic definition. The aim of the study is to focus on the particular unobtrusive features that are related to various types of understandings.

### 2.11.2 Communicative functions of micro-feedback

Since Duncan and Fiske (1977) and Oreström (1983), *micro-feedback* and its communicative functions have been widely discussed. Blöndal (2005) has identified a number of functions of communicative *feedback*, including *contact, perception/attention, carry-on signals*, and *understanding* (Blöndal, 2005; see Table 2.1). Classifications of micro-feedback, and other forms of feedback, have also been given by, for example, Nivre et al. (1992), Kopp, Grammer, Allwood, Ahlsén, Oberzaucher, and Koppensteiner (2007), and others.

Table 2.1. The types of linguistic and other communicative expressions of *feedback* (cited from Kopp et al., 2007, p. 257).

	<b>Bodily coordination</b>	<b>Facial expression, posture, prosody</b>	<b>Head gestures</b>	<b>Vocal verbal</b>
Awareness and control	Innate, automatic	Innate, potentially aware + controlled	Potentially/mostly aware + controlled	Potentially/mostly aware + controlled
Expression	Visible	Visible, audible	Visible	Audible
Type of function	C, P, E	C, P, U, E, A	C, P, U, E, A	C, P, U, E, A
Type of reception	Reactive	Reactive	Responsive	Responsive
Type of appraisal	Appraisal, evaluation	Appraisal, evaluation	Appraisal, evaluation	Appraisal, evaluation
Intentionality	Indicate	Indicate, display	Signal	Signal
Continuity	Analogue	Analogue, digital	Digital	Digital
Semiotic sign type	Index	Index, icon	Symbol	Symbol

C = Contact, P = Perception, U = Understanding, E = Emotion, A = Attitude

### 2.11.3 Definition

Based on the earlier definitions (Nivre et al., 1992; Allwood, 1993) and theoretical frameworks for classifying communicative *feedback* (Kopp et al., 2007; Grammer, Kopp, Allwood, & Ahlsén, 2008), the concept of *micro-feedback* can be further defined and operationalised.

*Micro-feedback* mainly refers to the unobtrusive vocal-verbal and gestural communicative expressions that are used to give and elicit information about the continuation of the interaction, the perception and understanding of the communicated information, and the attitudinal and emotional reactions to the perceived and understood information (see Nivre et al., 1992, pp. 1–5). In addition to these, *micro-feedback* in this study also has the following features: (1) *micro-feedback* items have no independent referential or semantic meaning but are very much dependent on the communication contexts; (2) they occur during or after the other speaker's talk, preferably at *response points* (Linell, 2011, pp. 266–270; see also *decision points* in Linell, 2013, pp. 57–89), usually at the beginning of a responsive communication contribution that includes both spoken utterances and gestural behaviours; (3) besides the basic communicative functions of *feedback* (i.e., CPUE/A) identified by Nivre et al. (1992) and Allwood (1993), *micro-feedback* also functions as a connector between the adjacent communication contributions; (4) they sometimes express positive and negative evaluative opinions, for example, agreement and disagreement,

besides the emotional and attitudinal reactions such as friendliness, surprise, hesitation, and so forth identified in Nivre et al. (1992).

Starting from the earlier definitions and theories of *feedback*, this thesis studies *micro-feedback* and its relation to understanding in social interaction. *Micro-feedback* and its features in terms of both modality (auditory and visual) and prosody will be investigated at the same time, something that has not been done in earlier research. An explorative approach is employed. This contributes to the conceptualisation and operationalisation of the concept of micro-feedback in communication.

#### 2.11.4 Modalities of micro-feedback

The notion, that micro-feedback has both vocal-verbal and gestural forms, has become commonly acknowledged in the relevant studies. In this thesis, *vocal-verbal* is used instead of *vocal* or *verbal*, because both *vocal* and *verbal* are too broad. For instance, *vocal* includes sounds that do not make sense (e.g., ordinary coughing or throat-clearing), and *verbal* includes sign language. Also, *gestural* is used instead of *nonverbal* because *nonverbal* is too broad in that it includes the vocal sounds that are not verbal (nonverbal) but make sense and have communicative functions (e.g., a kiss sound). *Gesture* and *gestural* in this thesis are defined to include all the body (or bodily) movements that have a communicative function (see Allwood, Cerrato, Jokinen, Navarretta, & Paggio, 2007; Allwood & Lu, 2011). By *communicative*, it is meant that the gestures play a role not only in maintaining contact and attention, but also in expressing understanding, activating, and illustrating the speech content. In this thesis, non-communicative or meaningless gazing at or other gaze movements, mouth movements, and blinks are not taken as anything more than natural human actions, which one person always does when he or she is awake. *Vocal-verbal* and *gestural micro-feedback* (see definitions later in this section) studied in this thesis is distinguished in terms of how it is produced (i.e., in relation to the auditory and the visual modality).

The term *modality* is used to refer to the *sensory modality* that was first introduced by Helmholtz in physics and was used to refer to a class of sensations (see Willis & Coggeshall, 2004, p. 1). Following Aristotle's classification of the five senses, that is, touch, vision, hearing, smell, and taste, the contemporary theory of sensory modality (or perception modality) focuses on haptic modality, visual modality, auditory modality, olfactory modality, and gustatory modality (see Kennedy & DeRuyter, 1991, pp. 123–190; Lyon, Nehaniv, & Cangelosi, 2007, p. 423). Although there is a coherent connection between the perception modalities and production modalities, for instance, gestural input is perceived through visual modality, vocal-verbal input

through auditory modality, tactile input through haptic modality, and specific chemical input through olfactory and gustatory modality, the term *modality* is used in this thesis from a perception-sensory perspective. This view has also been applied in education and learning studies, for example, Lindström, Marton, Emanuelsson, Lindahl, and Packendorff (2011). A given sensory modality would have a number of characteristics, such as quality, intensity, duration, and extension (Willis & Coggeshall, 2004, p. 1). As a tradition in the audio-visual communication studies, the auditory and visual modalities that are used in communication practices should always be given particular attention (see Lyon et al., 2007) rather than other sensory modalities. Therefore, in the present study, the focus is on auditory modality and visual modality.

Since perception sensory modality is considered to be different from production modality, terms like *bodily-visual modality* and *vocal-auditory modality* (see Zlatev, 2009, p. 158), *gestural modality* and *vocal-verbal modality* (see Lu & Allwood, 2011), and *auditory micro-feedback* and *visual micro-feedback* are not used in this thesis. One might want to differentiate perception from production of interaction in a more consistent way. That is, gestural micro-feedback is perceived by means of visual modality but produced through gestures; whereas, vocal-verbal micro-feedback is perceived through auditory modality but produced by vocal (articulatory) means. Therefore, in this thesis, the terms used about perceptual modalities are *auditory modality* and *visual modality*, and the terms about the production of micro-feedback are *vocal-verbal micro-feedback*, *gestural micro-feedback*, *unimodal vocal-verbal micro-feedback*, *unimodal gestural micro-feedback*, and *multimodal micro-feedback*.

*Unimodal micro-feedback*, in the present study, refers to the micro-feedback that involves only one sensory modality (i.e., auditory or visual modality). It includes *unimodal vocal-verbal micro-feedback*<sup>27</sup> such as *yeah*, *okay okay*, and *that's right* and *unimodal gestural micro-feedback* such as head nod, smile, shrug, and eyebrow frown (i.e., the communicative micro-feedback information that involves only auditory or visual modality, respectively). As mentioned earlier, non-communicative bodily behaviours are not taken into account in this study. Communicative gestures through visual modality are not presupposed to exist all the time in spontaneous communication. Unimodal vocal-verbal micro-feedback can still be identified when it occurs without any communicative gesture but with some non-communicative ones, for example, when one interlocutor says *yes you are right* while he is scratching his head and gazing at the other interlocutor. *Multimodal micro-feedback* refers to the micro-

---

<sup>27</sup> Details of the classifications of *vocal-verbal micro-feedback* and *gestural micro-feedback* are presented in Section 4.1.9.

feedback that involves more than one sensory modality, for example, both the auditory modality and the visual modality. In a communication situation with both video and audio channels, one micro-feedback item can consist of both vocal-verbal and gestural micro-feedback components. This thesis will contribute to previous research on whether and how very different communication channels impact on micro-feedback and thus understanding.

## **2.12 Summary: exploring understanding by means of micro-feedback and meaning repair**

As discussed earlier, empirical understanding in human interaction is difficult to operationalise. In the research on understanding in interaction, qualitative studies are more common than quantitative ones. Also, a number of researchers, for example, Schegloff (1992), Mustajoki (2012), Verdonik (2010), Weigand (2010), Wierzbicka (2010), and Sarangi (1994), with different focuses and from various perspectives, have pointed out that counting and accounting for understanding problems and miscommunications is problematic.

Micro-feedback, as discussed, especially its most minimal forms, has usually been regarded as continuers or go-ahead signals (Schegloff, 1982). However, I will study these forms of micro-feedback primarily as signals of understanding. This is justified by the assumption that there must be forms of (shallow) understandings underlying the giving of continuers.

Because there have been mostly qualitative studies of understanding in conversation, this particular study will investigate understanding in both qualitative and quantitative methods. Quantifying understanding cases is of interest primarily for comparing the occurrences of different understandings across various communication contexts. Accordingly, it will be possible to present a general view of what happens to understanding in real-time communication, in particular in these two social activities focused on.

The occurrences of understanding and understanding problems will be quantified according to the frequency of micro-feedback expressions and sometimes of other types of responsive interactions, primarily, meaning repair in the studies in this thesis. Accordingly, studying understanding with analytical focuses on micro-feedback and meaning repair may provide an opportunity to measure and compare understanding in interactions. Thus, a further empirically based understanding of how understanding problems occur and how they are coped with in interaction can be obtained, and comparisons of them between different communication contexts can be attempted.



# 3

## Research review

This chapter presents the research review of understanding in interaction, understanding in intercultural communication, ICT influence on communication and understanding, and earlier studies of micro-feedback and prosody. It will show what has or has not been done in earlier research and what will be done in this thesis, with the aim of exploring the concepts of micro-feedback and understanding and the relation between them.

### 3.1 Earlier research on understanding in interaction

Most of the earlier studies on understanding in communication are theoretical, and there is in general a substantial lack of empirical research. A comparison of the number of studies on misunderstanding shows there are fewer relating to (sufficient) understanding and non-understanding. Nevertheless, a few of these theoretical studies will be presented in this section, in order to help understand the subject (i.e., understanding in communication) more deeply.

#### 3.1.1 Understanding

Earlier studies of understanding in conversation (e.g., Zaefferer, 1977; Thomas, 1983; Van Dijk & Kintsch, 1983; Bazzanella & Damiano, 1999; Weigand, 1999; Dascal, 1999; Verdonik, 2010) have mostly focused on vocal-verbal aspects instead of gestural behaviours. This is possibly because of the tradition of conversation analysis, or

simply because vocal-verbal behaviours reveal more information concerning understanding than gestural ones.

According to Schul and Lamb (1982), although having been often neglected, gestural aspects add value to the vocal-verbal aspects in information sharing and communication understanding. However, they did not explain further what and how much this value was. Gumperz (1982) has acknowledged that society and culture influence both *verbal* and *non-verbal communication* (p. 131). Borod (1993) claimed that the visual gestural behaviours may be just as important as the auditory vocal-verbal ones in social interaction. Later, Navarretta and Paggio (2013) found in an empirical study that gestural behaviours such as head movements and facial expressions play a large role in communication. Gesture not only adds information to the vocal word, thus having a contingent meaning together, it also contradicts or changes the semantic meaning of the vocal word.

In recent years, gestural behaviours have attracted increasing attention in language and communication research. A study of gestural behaviours would arguably extend and deepen our understanding of the phenomenon of understanding in interaction. Indeed, the role of the body in conveying understanding has been pointed out by Hindmarsh, Reynolds, and Dunne (2011): “the body is a central resource to exhibit and to assess understanding in real-time” (p. 501). There are isolated examples of analysis of gestural expression in understanding (e.g., Weigand, 1999), but there have been few systematic studies of gestural and multimodal aspects of understanding.

Mondada (2011, p. 545) noted that “gesture and other multimodal conducts are often observed by focusing on speakers—the speaker being the one who gesticulates during his talk—multimodal conducts of recipients have been less studied”. The present study aims to remedy this situation by analysing how interlocutors convey understanding through micro-feedback, bodily and verbally, in both unimodal and multimodal forms.

### 3.1.2 Misunderstanding

Since the 1970s, misunderstanding<sup>28</sup> has become a favourite topic in sociolinguistics and discourse analysis (Zaefferer, 1977; Gumperz, 1982; Schegloff, 1987; Weigand, 1999; Verdonik, 2010). The greatest interest in misunderstanding was shown in 1999

---

<sup>28</sup> “Misunderstanding should not be confused with *misconception*. A misconception is an error in the prior knowledge of a participant” (Hirst et al., 1994, p. 214). “A misunderstanding is the communicatee’s choice of an interpretation for an utterance which is not the one intended by the communicator” (Sayer, 2013, p. 738).

in a special issue of the *Journal of Pragmatics* (vol. 31, 1999). However, according to Verdonik (2010), there have been few studies on misunderstanding in recent years. *Misunderstanding* has been addressed in different terms, such as *pragmatic communication failure* (Thomas, 1983) and *reduced understanding* (Verdonik, 2010). They are applied to a variety of cases and perspectives within varying methodological frameworks.

According to Schegloff (1987) and Weigand (1999), our languages are built for effortless understanding and communication, therefore it is usually presumed that the standard case of language use presupposes co-membership and deals with communication between members of the same community and the same cultural world (cf. Weigand, 1999, p. 764). However, Gumperz (1982) pointed out that the *socio-cultural conventions* (p. 131) influenced all levels of speech production and interpretation. *Intercultural* (i.e., between two cultures) and *cross-cultural* (i.e., among three cultures or more) *communication* problems can be observed not only between speakers of different languages but also between speakers of different varieties of the same language. Therefore, although Schegloff (1987) and Weigand's (1999) harmonious model as somewhere to begin with can be a way to approach understanding, it is too ideal to presuppose the person that one is communicating with has the same culture or shares the same knowledge. Perhaps, that is why, apart from misunderstandings, at times situations of uncertainty and vagueness occur in communication, for example, when people seek the right way to communicate with one another by referring to their own culture or knowledge.

Verdonik (2010) noticed that the source of vagueness and ambiguity may thereafter lead to a *reduced understanding* (p. 1367) or misunderstanding. Nevertheless, communication in itself allows for, contains, and fosters a variety of convergences and divergences between communicators, which can be global or local, collective or individual, and common or special. "We must abandon the existing views of communication which draw a basic distinction between cultural or social knowledge on the one hand and linguistic signalling processes on the other" (Gumperz, 1982, p. 186). Instead, the interactional approach of contextualisation should be taken into account when looking at understanding and misunderstanding in communication.

As regards the classification of misunderstanding, de Souza and Danilo (1985) pointed out that *planned misunderstanding* occurs when people who know how misunderstanding works deliberately communicate in a way that leads to misunderstanding, although it may not be common in normal social interaction. Also, according to Blum-Kulka and Weizman (1988) and Weigand (1999), misunderstanding is not normally detected by the interlocutor. Thus, an *unresolved misunderstanding* occurs. This is usually in a connection with indirect speech acts (for more details, see

Blum-Kulka & Weizman, 1988), since indirectness leaves things more open and unresolved. Moreover, Dascal (1985) proposed *preventing misunderstanding*, which in contrast to *planned misunderstanding* means the speaker tries to secure understanding and thus *prevent misunderstanding*. The reason this categorisation of misunderstanding is not used in the study is that misunderstanding normally occurs without being intended or planned by any speaker and that if the misunderstanding is prevented then it does not exist (it is not a misunderstanding) any more. However, these theories all seem to suggest that misunderstanding can be explicit or implicit, depending very much on the interaction context and situation.

Zaefferer (1977) analysed misunderstanding and made a distinction between *misunderstanding* and *non-understanding*. *Misunderstanding* is similar to *understanding*, which involves attempts at making sense, whereas *non-understanding* does not. This is an interesting theory, which suggests at least three categorisations of understanding: understanding, misunderstanding, and non-understanding. However, it is not effective to have one category of the concept of understanding also called understanding. Also, does non-understanding involve no sense-making at all or just no achieved (but failed) sense-making? Zaefferer (1977) mentioned a few classes of misunderstanding such as *entirely correct partial understanding* (p. 344), *partial understanding* (p. 344), *misunderstanding in a wide sense* (p. 344), and *misunderstanding in a narrow sense* (p. 344), which are highly complex and difficult to distinguish. Zaefferer (1977) stressed that various types of misunderstandings can influence the phonological, syntactic, semantic, or situational levels as well as the illocutionary force and the propositional content. However, he did not elaborate on the relationship between grammar and context. That is, contextualisation is neglected in his theory of misunderstanding. Bazzanella and Damiano (1999) had a similar taxonomy, which consists of the phonetic, syntactic, lexical, semantic, and pragmatic levels. Their aim was to study the triggers and the handling of misunderstanding in conversations. Similar to Zaefferer (1977), Bazzanella and Damiano (1999) also emphasised the difference between *understanding* and *misunderstanding* on the one hand and *non-understanding* on the other. They argued that understanding and misunderstanding should be seen as two ends of a continuum rather than two discrete cases. However, as discussed earlier, some misunderstandings are constructive, but some are not. Thus, misunderstanding does not necessarily or always lead to any so-called “understanding”, and it is not obvious why understanding and misunderstanding should be treated together as one separate category. Although Zaefferer’s (1977) and Bazzanella and Damiano’s (1999) approaches are analytically interesting when stud-

ying understanding, none of them has provided a concrete criterion for how to identify and operationalise understanding, misunderstanding, and non-understanding as classified by them.

Thomas (1983) considered misunderstanding to be *pragmatic communication failure* and classified it into two levels. On the first level, the interlocutor fails to understand the proposition correctly; and on the second level, the interlocutor fails to understand the intended pragmatic force, which includes *pragmalinguistic failure* and *sociopragmatic failure*. *Sociopragmatic failure* in particular, as pointed out by Thomas (1983), typically occurs in intercultural communication contexts and mainly refers to the social conditions that are placed on language use in different cultures. For example, an Indian head shake and nod can mean “yes” and “no”, respectively, which is just the opposite to what the Swedish and Chinese speakers normally do. Thus, how to investigate the use of language (or a certain linguistic means) while taking into account the cultural and social backgrounds seems important. In this thesis, although neither the cultural nor the individual or gender differences are within my research scope, attention will be still paid to controlling the reliability of the interpretation and understanding of the empirical interaction data. For instance, both Chinese and Swedish transcribers and annotators are involved in transcribing and annotating each interaction studied (see Sections 4.1.10 and 4.2.8).

### 3.1.3 Non-understanding

Dascal and Berrenstein (1987) approached understanding mainly by distinguishing *comprehending* and *grasping*. According to Dascal and Berrenstein, *comprehending* refers to being able to understand the pragmatic level, such as the meanings of the sentence, the utterance, and the speaker; and *grasping* means being able to detect what can and cannot be said in a given situation, for instance, determining what rules (of social interaction) should be followed. Different types of misunderstandings and non-understandings correspond to each of them. As Weigand (1999) questioned, however, is there any rule to be determined in social interaction? If any, can the interlocutors really determine it? For instance, in this dialogue excerpt in Weigand (1999), “A: John comes back tomorrow. B: What did you say? I don’t understand, the television is too loud” (p. 770), there is obviously a case of non-understanding going on. However, what rule is used in this communication failure of non-understanding in particular? How is the rule determined? Which part of this dialogue is the so-called *comprehending* and which is *grasping*? Apparently, Dascal and Berrenstein’s (1987) theory is not able to solve these empirical problems.

Van Dijk and Kintsch (1983) viewed understanding as a cognitive process of discourse comprehension. With this theory, Weigand (1999) proposed a model of dialogic action game (see Figure 2.1), which concentrates on social interaction. She made a number of claims as follows. First, language cannot be separated from human beings. Secondly, human beings are at the centre of dialogic action and have activities with different purposes and needs. Thirdly, knowledge is socio-culturally and personally variable. Individual differences, for example, in cognitive backgrounds and personal experiences inevitably lead to consequences of different understandings, especially misunderstandings and non-understandings. According to Verdonik (2010), understanding is viewed as a *creation of meaning*, which depends on factors related to the person who understands as well as the social and contextual factors. This view is very close to the interactional perspective applied in the present thesis. As Verdonik (2010) suggested, understanding was complexity oriented, building from the most local (or individual) to the most global (or collective), but continuously looking back at the local (or individual). If the knowledge of relevance (e.g., common ground when it comes to culture, language, topic, and subject) is not sufficient, the interlocutors will not be able to share the same sense-making and will have problems achieving mutual understanding. Thus, misunderstanding and non-understanding occur.

## **3.2 Research on understanding in intercultural communication**

Understanding in intercultural communication is attended to in this thesis, because communication and culture are closely linked. As Neuliep (2016) has said, “Culture shapes communication, and communication is culture bound” (p. 8). Culture is an accumulated pattern of values, beliefs, and behaviours shared by an identifiable or particular group of people with a common history and verbal and non-verbal symbol systems that are learned and not given by nature (cf. Allwood, 2015).

### **3.2.1 On intercultural communication**

When people from different cultures come together and communicate with one another, intercultural communication takes place. With reference to Ting-Toomey (1999), Lustig and Koester (2003), Holliday, Hyde, and Kullman (2004), Lindström (2008), and Allwood (2015), the term *intercultural communication* will be used about the communication between people from different cultures, that is, groups of people who come from different linguistic, ethnic, and social backgrounds and who may have different patterns of thought, belief, values, norms, and social behaviour and do

not have any dominant culture in common. The academic discipline of intercultural communication can be traced back to Hall's (1959) *The Silent Language*. "Hall is generally recognized as the founder of the academic discipline we call intercultural communication" (Neuliep, 2016, p. 14), and the anthropological approach is the historical origin. It is believed that culture is an adaptation to and a distinctive product of a unique set of historical, social conventional, and environmental conditions. "As these conditions vary, cultures vary accordingly" (Neuliep, 2016, p. 14). There is no correct or standard culture, and there is hardly a stereotype of culture. People are often as unaware of their own cultural ways of living and communicating as they are of pursuing understanding. However, intercultural communication "has always been and probably will remain an important precondition of human co-existence on earth" (Allwood, 2015, p. 1). This pinpoints the importance of intercultural communication in social interaction.

As Holliday, Kullman, and Hyde (2016) have said, "understanding in intercultural communication should grow from an understanding of people, culture, and society generally" (p. 1). People all over the world need effective and competent intercultural communication (Neuliep, 2016). More importantly, "the need ... is felt intra-personally, within our own personal, social, and professional lives and relationships" (Neuliep, 2016, p. 5) and is also felt interpersonally, between different persons of various linguistic, ethnic, and social backgrounds.

### 3.2.2 On understanding in intercultural communication

Culture and cultural differences affect communication and increase understanding problems (Gogan, Popescu, & Duran, 2014). Intercultural communication likely has higher risks of lack of understanding and misunderstanding (see also Allwood, 2015; Lindström, 2008).

Gumperz (1982), Pride (1985), Tannen (1990), Smith (2004), Eliot (2010), and Samovar et al. (2012) have pointed out that socio-cultural conventions affect all levels of speech production and interpretation. When people who have different cultural and language backgrounds enter into a joint intercultural communication activity, it is very likely that they have more problems and difficulties in understanding and sense-making than when they communicate with people who have the same relevant backgrounds.

Neuliep (2016) argues that the most obvious difference between two cultures is language. That is, "differences in language and culture quite naturally lead to communication difficulties" (Weigand, 1999, p. 764). Gogan et al. (2014) have stressed that English as the lingua franca has become a common tool for intercultural communication, and that misunderstanding often happens because of the pronunciation

and grammar errors of the English language users who have other native languages. Considering that “the English language is full of meaning nuances; a word may have multiple meanings based upon the context that it is used” (Gogan et al., 2014, p. 373), it is very common that people encounter understanding difficulties more frequently in intercultural communications than in mono-cultural settings, that is, communication within a single social or ethnic group. These understanding problems lead to miscommunication, poor leadership and ineffective teamwork, and sometimes meaningless conflicts to various extents.

A number of language and communication researchers have made efforts to give advice on improving understanding in intercultural communications. For example, Draghici (2007), Zofi (2011), and Popescu, Aldea, and Draghici (2012) have stressed that raising questions separately (i.e., letting the other answer one question at a time), taking turns (i.e., making a point and then listening to the response), checking meanings and being active in verifying the perceived information, giving feedback, and keeping an eye on the mood can be important to help understanding and avoid misunderstanding. However, not much empirical research has been done on analysing the actual occurrence of understanding problems and how they are actually coped with in intercultural communication. Both Study 1 and Study 2 in this thesis are based on empirical data of intercultural communication.

### **3.3 Influence of ICT on communication and understanding: VMC vs. FTF**

As information and communications technology (henceforth ICT) develops, computer mediated communication (henceforth CMC) has been experienced as fast in connectivity, rapid and secure in exchange of information, and easy and relaxed in engagement to interaction (Garner & Buckner, 2013; Wright & Webb, 2011; Stacks & Salwen, 2009). Irrespective of different modes of ICT and CMC (e.g., text-based or oral, synchronous or asynchronous), people have increasingly indulged in having daily contacts, business communication, and learning activities using CMC (Ziegler, 2016; AbuSeileek & Qatawneh, 2013).

#### **3.3.1 Computer mediated communication (CMC)**

Extensive studies of ICT influence on life have been carried out, although most often with a focus on communication coordination and learning effects in the disciplines of communication and education. For instance, Abrams (2008) pointed out that CMC helped learners to interact and negotiate the learning content actively, provided rich opportunities for them to recognise and adapt to diverse intellectual and

interactional patterns, and also developed their learning ability, linguistic knowledge, and communicative competence. AbuSeileek and Qatawneh (2013) studied the English language learners' use of discourse functions in synchronous CMC (audio chat) and found that the question types and strategies were usually short, clear, and unambiguous, and they took this as confirmation of the effectiveness of language use in CMC. Sins et al. (2011) researched ICT effects on high school students' performance in a collaborative modeling task and found that the collaboration results in FTF (face-to-face chat) and CMC (text chat) were equally good and the CMC group was more effective at reasoning and discussing than the FTF group. With a similar topic but a different context, Hatem, Kwan, and Miles (2012) compared the effectiveness of CMC (video conferencing) collaboration in the construction industry and found that people collaborated in CMC as effectively as in FTF and sometimes even slightly more effectively. Most of these earlier research into CMC seems to focus on studying the effectiveness and the result of team (working or learning) collaboration. More or less, one can conclude that CMC collaboration is as effective and good as, and sometimes more effective and better than, FTF.

### 3.3.2 Video-mediated communication (VMC)

One popular mode of CMC, video-mediated communication (henceforth VMC), with its high image resolution and voice quality, fast signal transmission and alignment, and temporal (time) and spatial (logistics) effectiveness, is more commonly applied in various CMC activities, such as video conferencing and web video tutorials. Besides numerous advantages of VMC, the notion of "social presence" or the lack of it in VMC has been highlighted since the 1970's (Short, Williams, & Christie, 1976; cf. Anderson, 2006). Compared to FTF, the communication VMC's channel is relatively narrow and media richness is comparatively low (Sins et al., 2011). VMC somewhat restricts the exchange of auditory and visual communication cues, such as prosody and gesture, which normally help people to regulate interaction, perceive, express, and comprehend information (Driskell & Radtke, 2003), and monitor feedback from others (Straus, 1997). Even when VMC has good sound and image quality, there may be difficulties in grounding (cf. Anderson, 2006), because not all the sources of information available in FTF are transmitted. The explicit communicative coordination micro-feedback has often turned out to be unclear and infrequent (Paul, Seetharaman, Samarah, & Mykytyn, 2004; Friedman & Currall, 2003), even though it is very critical for communication understanding (McIntyre & Salas, 1995). Most of these earlier studies of VMC seem to focus on studying the interlocutors' linguistic behaviours rather than the effectiveness or result of the interlocutors' team collaboration, as different from the earlier research on other forms of CMC.

### 3.3.3 Comparing to face-to-face communication (FTF)

Although Newlands, Anderson, and Mullin (2003) have shown that people adapt to the communication media and adjust the way they collaborate in exchanging information and achieving mutual understanding, it is still found that certain forms and patterns of communication behaviours are different between VMC and FTF. For example, compared to face-to-face (henceforth FTF) communication, interlocutors in VMC have lower interdependency (Stone & Posey, 2008), lower cohesion and less conformity (Kiesler & Sproull, 1992), and more difficulties in achieving communication goals, or have to work harder to achieve them (Whittaker, 2003). Interlocutors in VMC are more unaware of each other's non-verbal communication behaviours (Stone & Posey, 2008), they often have more trouble understanding others' responses (Kiesler & Sproull, 1992), and they are more likely to experience ambiguity and problems in understanding (Shin, 2005). Communication technologies often lead to difficulties in turn taking, and speakers in VMC are slower to detect and correct misunderstandings (Thompson & Coover, 2006) and more constrained in achieving common ground and achieving understanding (Olson & Olson, 2000; Clark, 1996; Clark & Brennan, 1991). Most of these studies focus on investigating the communicative and interactive behaviours in VMC versus FTF, and most of these researchers seem to agree that the limited linguistic channels in VMC more likely result in miscommunication and understanding problems than in FTF.

However, a few researchers have found that mediating technology has little effect on interacting and understanding (e.g., Anderson, 2006). These studies often focus on investigating the outcome of team problem-solving collaboration in VMC versus FTF, which allow clear measurement of who knew, understood, and did what and also how and to what extent the problems were solved. For example, Anderson (2006) analysed participants' laboratory map task-solving interactions in FTF and VMC and found that VMC has little effect on achieving understanding between the participants and the content of the interactions. This makes VMC an interesting area in which to study understanding. Little research has been done on what understanding difficulties and problems are present in VMC and FTF interactions and how they are coped with in interaction. Understanding and understanding problems with richer empirical data from both FTF and VMC interactions will be investigated in Study 2 in this thesis.

### 3.4 Previous studies of micro-feedback

In the field of linguistics and communication, there have been a number of studies on micro-feedback, although with different research aims and different terminologies (as presented in Chapter 2). The term *micro-feedback* instead of other variations will be used in the following sections.

#### 3.4.1 Functions of micro-feedback

As presented earlier in the theoretical frameworks, micro-feedback has communicative functions of signalling contact, perception, understanding, and emotional and attitudinal reactions. So, how do these functions work in practice? Cutrone (2010) conducted a study of micro-feedback in intercultural conversations between native English teachers and Japanese second language learners of English. Cutrone found that micro-feedback was important in intercultural conversations and people from different countries and cultures had different conventions for using and interpreting micro-feedback. Cutrone (2010) said that when a Japanese speaker says *yeah* or *yes* it does not necessarily mean he or she understands and agrees. Instead, from the accompanied intonation and gesture, these micro-feedback words can have a variety of meanings (perhaps combined with different emotions and attitudes). In many situations, *yeah* or *yes* can mean the speaker does not understand or agree. It does not necessarily signal understanding and agreement or acceptance at the same time. As in Cutrone's (2010, p. 34) example:

Sometimes these misunderstandings can have dire consequences as was the case in the Hitachi-Mitsubishi trial (Japan Times 1983). One of the defendants in the case, Mr. Ishida of Mitsubishi claimed that he had not agreed with the FBI undercover agents when they told him he had stolen some information documents. His defense counselor argued that Mr. Ishida's responses of *yeah* and *uhuh* were not to show agreement, but rather to indicate he was listening and understood what was going on and to allow the other person to continue.

Cutrone's (2010) study not only emphasises the socio-cultural perspective on studying micro-feedback, it also suggests that besides the vocal-verbal content of micro-feedback there are also other factors such as gesture and prosody that interfere with its meaning and functions in social interaction.

#### 3.4.2 Modality of micro-feedback

Kopp et al. (2007) and Grammer et al. (2008) have recognised that the micro-feedback system in human FTF interactions comprise much more than spoken words in

their framework. For example, “the interlocutors incessantly coordinate and exchange *feedback* information by *nonverbal* means” (Kopp et al., 2007, p. 256) like posture, head movement, facial expression, and so on. Grammer et al. (2008) used this framework as a theoretical model for virtual agent communication, and they further stressed that the micro-feedback system involves two primary types of content, that is, visible gestures and audible vocal words. Although the researchers did not mention how the audible and the visible micro-feedback expressions are coordinated (e.g., in a complementary way), their framework for analysing micro-feedback has still been widely adopted in many communication studies. In my study, I will focus on the relations between vocal-verbal and gestural micro-feedback.

### Vocal-verbal micro-feedback research

There are a number of studies that made claims about the most frequently used vocal-verbal micro-feedback expressions in speech. For instance, Jurafsky, Shriberg, Fox, and Curl (1998) found that the most frequent micro-feedback expressions were *yeah*, *uh-huh*, *hm*, *right*, and *okay*. Kopp et al. (2007) recognised *yes*, *no*, and *m* as among the most frequent micro-feedback words in spoken language. Also, Lu and Allwood (2011) confirmed *yeah*, *okay*, *m* as the top expressions in their study.

There are certainly more studies than these, but it is not possible to present them all here. For example, Blöndal (2005) studied micro-feedback in Icelandic conversational storytelling, which was based on 30 conversations (approximately 20 hours) from the Icelandic Corpus of Spoken Language (ISTAL). The purpose of the study was to determine the distribution of micro-feedback (primarily vocal-verbal micro-feedback) in various parts of the stories, and also to explore its forms and functions. Besides the finding of the most frequently used vocal-verbal micro-feedback expressions, Blöndal found that both the listener and the storyteller had some kind of anticipation of when the vocal-verbal micro-feedback should be delivered. For instance, as soon as the listener has heard a preposition in association with what he anticipates (i.e., at what Linell (2011, pp. 266–270) calls “response points”), micro-feedback words such as *mhm*, *okay* are articulated, which means “I know, I understand, I am with what you are saying”. Also, when the listener utters, for example, *mhm* or *ja* in a place that is inappropriate according to the storyteller’s anticipation, the storyteller sometimes begins to repeat what he or she has just said before and then continues to tell the story. This finding is in line with Ward (1996) and Ward and Tsukahara’s (2000) ideas (for more details, see Section 3.5.2). Although Blöndal did not study in detail how micro-feedback expressions, vocal-verbal ones in particular in her study, were related to understanding, she still pointed out that many *ja* and *mm* in the data signified understanding rather than just filling in the missing word. Besides these,

Blöndal investigated laughter separately and claimed that laughter was very often used to show that the story was understood and appreciated by the story listener. This may shed light on the necessity of investigating the relation between micro-feedback and understanding as well. Blöndal said that laughter did not always have the function of micro-feedback, however, she did not present what other communicative functions laughter might have in the study. This is brought up, because firstly it is assumed that laughter (or chuckle) in interaction is always micro-feedback, at least in normal human interactions. Secondly, it should be stressed that in the present study laughter (or chuckle) is treated as a multimodal micro-feedback expression, which involves both auditory and visual modalities rather than vocal-verbal only or primarily. Although laughter (or chuckle) is not the focus of the present study, it is an interesting topic and could be the subject of research in the future.

### Gestural micro-feedback studies

With respect to *gestural micro-feedback*, Cerrato (2005) conducted a study of the gestural characteristics of *m*-like sounds in Swedish, based on two video-recorded travel agency conversations. She investigated the relationship between the *m*-like sounds, which are mostly micro-feedback expressions, and their accompanying gestures. Cerrato found that 41% of all the analysed *m*-like sounds were accompanied by a gesture and the most common ones were head movements such as down-nods and up-nods. Besides this, she proposed a few types of relationships between gesture and its accompanying speech. That is, gesture can have a neutral or non-marked function, or can add, emphasise, weaken or contradict speech. Furthermore, Cerrato (2005) found that gestures were often made to emphasise some information, which was also focused on or emphasised by the prosody of speech. This suggests that gesture and speech may be interdependent and that both play a complementary role to each other in human interaction. Cerrato's (2005) study shed light on the importance of gestural micro-feedback, in particular head movements, in social interactions.

Navarretta and Paggio (2010) analysed eleven map-task dialogues between Danish speakers from the Danish Phonetically Annotated Spontaneous Speech Corpus. Among other things, Navarretta and Paggio found that gesture in micro-feedback had a strong effect on the classification of dialogue acts. Later, Paggio and Navarretta (2011) carried out another study of gestural micro-feedback behaviour in a corpus of twelve dyadic Danish first encounter dialogues, with a focus on head movements as signals of micro-feedback in interaction. As a result, Paggio and Navarretta found that the most frequently used micro-feedback gestures were head movements, especially single nod and repeated nods, and that about 67% of all the head movements were used to express micro-feedback. Furthermore, they also pointed out that it was

important to analyse how other gestural behaviours apart from head movements, such as facial expressions, are related to micro-feedback. Although Paggio and Navarretta (2011) stated that “participants... that cannot see each other need to check mutual understanding and grounding by using feedback words” (p. 37), they did not pay enough attention to the vocal-verbal micro-feedback in their study.

In this thesis, both vocal-verbal and gestural micro-feedback, which not only include the head movements but also all the other communicative gestures of relevance, will be taken into account. This is because the vocal-verbal micro-feedback not only communicates the verbal message. For instance, one can sense the other’s hesitation (e.g., in understanding) by hearing a long, slow, and reluctant *yeah* or *okay*. Gestures usually provide information that complements the speech content in FTF interactions. For example, one can perceive the other’s sufficient understanding by seeing a big and firm nod. Thus, both vocal-verbal and gestural micro-feedback expressions play important roles in spontaneous communication. This thesis will investigate how understanding is in particular signalled or conveyed through micro-feedback in both auditory and visual modalities.

### Multimodal micro-feedback literature

As regards *multimodal micro-feedback*, Merola and Poggi (2004) studied the multimodality in teacher communication, with a focus on analysing both the affective and interactive aspects of vocal-verbally and gesturally communicated information and its cognitive effects. The analysis was based on video-recordings of four female teachers in three different classes of 7-year old pupils. Among other findings, the relationship between the meaning of the gesture and that of the accompanied vocal word was identified. According to Merola and Poggi (2004, p. 107), the relationship can be repetitive if both modalities provide the same information, additive if one provides additional but congruent information, substitutive if one modality provides some information that is not given by the other, contradictory if the information expressed by one modality is incompatible with the information expressed by the other, or indifferent if one modality forms part of a different communicative plan. Merola and Poggi (2004) focused on the relationship between the vocal words and the gestures that comprise one multimodal unit. However, they did not discuss which specific gesture and vocal word usually work together and comprise such a multimodal unit in their study.

Paggio and Navarretta (2013) further investigated the relationship between gestural behaviour and vocal-verbal speech with regard to the expression of micro-feedback in particular. Their study was based on two Danish data corpora consisting of map-task dialogues and first encounter conversations. As a result, they found that

the gestures and the speech disambiguate each other in the machine learning process. They also found that head movements and to a lesser extent facial expressions were important indicators of micro-feedback. That is, both head movements and facial expressions are often combined with micro-feedback speech. Besides these, Paggio and Navarretta (2013) claimed that most of the multimodal micro-feedback expressions consisted of speech (*yes* and *no* particularly) and head movements or facial expressions often had a dialogue act of acceptance and an emotional function of agreement. Although multimodal micro-feedback has been more or less addressed by Paggio and Navarretta (2013), it was in particular focused on the micro-feedback items that consist of speech *yes* or *no* and head movements or facial expressions. Multimodal micro-feedback involving other vocal-verbal and gestural components than *yes* or *no* plus head movements or facial expressions still requires further research. In this thesis, all the multimodal micro-feedback expressions that occur in the empirical data will be taken into account.

### 3.5 Research literature on prosody

In this section, a number of studies of prosody in general in the field of linguistics and communication will be presented first. Then, some prosodic studies that focus specifically on vocal-verbal micro-feedback will be presented.

#### 3.5.1 Studies of prosody in linguistics and communication in general

Back in the 1960s, Abercrombie (1965, p. 4) addressed the importance of prosody in conversation:

If you are reading aloud a piece of written prose, you infer from the text what intonations you ought to use, even if, as is almost always the case, you have a choice. The intonation, in other words, adds little information. But if you try to read aloud a piece of written conversation, you can't tell what the intonations should be—or rather what they actually were. Here the intonations contribute more independently to the meaning.

In the 1990s, Couper-Kuhlen and Selting (1996) pointed out that prosody as a field had been left unexplored by modern linguists, in that “only a few scholars at the most have considered prosody, intonation in particular, worthy of their attention” (p. 11). Intonation has thus begun to be seen as “a powerful means of creating interactional meaning, in alignment or non-alignment with verbal forms” (Couper-Kuhlen & Selting, 1996, p. 22). Couper-Kuhlen and Selting proposed an interactional perspective

to approach prosody, and they suggested that prosodic features can be treated as interlocutors' devices designed for the organisation and management of talk in the social interaction.

In another study, Couper-Kuhlen (1996) pointed out that there were two ways of repeating the pitch register of another interlocutor, that is, *relatively repeating* (i.e., *quotation*) and *absolutely repeating* (i.e., *mimicry*). Besides this, Couper-Kuhlen (1996) also analysed the relationship between prosodic repetition and verbal repetition from a contextualised and interactional perspective. She claimed that "lexico-syntactic or verbal repetition may occur with differing prosodic or non-verbal configuration, and prosodic repetition may occur with differing lexico-syntactic carriers" (Couper-Kuhlen, 1996, p. 367). She pointed out that the prosodic repetition included the repetition or copying of syllable loudness, syllable duration, and syllable pitch. As Couper-Kuhlen and Selting (1996, p. 48) stressed, "the study of prosody should be based on empirical data from natural interaction". Couper-Kuhlen (1996) studied a number of conversations from a corpus of British phone-ins, and concluded that relative and absolute pitch registers were not arbitrary but depend on the natural voice ranges of the speakers involved. This suggests that individuality plays an important role in studying prosody. Although neither prosodic repetition nor verbal repetition is the focus of the present study, Couper-Kuhlen's (1996) study sheds light on my thesis in the following ways. First, prosody is a contextualised feature of vocal-verbal speech, and it plays a role as important as gesture and speech in social interaction. Second, prosody is coordinated with the vocal-verbal speech to communicate information. The meaning of the communicated information can change from one sense to another by means of different prosodic features. There is a complementary relationship between prosody and speech content. Third, individual differences are prominent when studying pitch features in communication. Fourth, loudness (intensity), duration, and pitch are the primary elements of prosody in interaction, which is in line with what Pollack, Rubenstein, and Horowitz (1960) and Crystal (1969) proposed.

Since the 1990s, prosody in communication has attracted more attention. Schegloff (1998) pointed out that prosodic features of speech had certain associations with communication functions and *conversational actions* (p. 235), for example, turn management. Schegloff examined three episodes of two telephone conversations. He found that the recipients (as Schegloff called them) may start a next turn after the pitch peak, which projects a completion of the syntactic or semantic construction. Also, the decelerating (as Schegloff called it) prosodic pattern may also project the possibility of a completion of the syntactic or semantic construction. Schegloff stressed the importance of prosody in conversation by saying that prosodic features

of speech in context can suggest as much information as syntax and semantics usually do. However, Schegloff focused more on pitch contour than pitch, or duration, or other prosodic features.

Nöth et al. (2002) investigated how prosodic information can be used in automatic dialogue systems. They dealt with the prosodic properties, that is, pitch, loudness, and length of the perceived speech, and they measured the acoustic correlates of perception, that is,  $F_0$  (i.e., pitch in the present study), energy (or intensity), and duration. This approach to studying prosody is in line with Pollack et al. (1960), Crystal (1969), as well as Couper-Kuhlen (1996). Although their study focused on the technical automatic dialogue recognition system rather than real human conversation, Nöth et al. acknowledged that syntax and semantics were traditionally, and in fact, the mediator between prosody and dialogue act (such as those of greeting, suggesting, requesting, and accepting). Although Nöth et al. tried to integrate the paradigmatic, the syntagmatic, and the pragmatic approaches to studying prosody in automatic dialogue system, they concluded that such a highly sophisticated method can only correspond closely with the strategies of human beings in human-human communication but cannot present them fully. Accordingly, prosody in the empirical data from spontaneous human conversations will be studied in this thesis.

In recent decades, prosody has been studied in relation to emotions and attitudes, although it is extremely difficult for researchers to reach a consensus on the taxonomy of emotions and attitudes (see Couper-Kuhlen, 2009). For instance, even the universality of Ekman's (1992) basic emotions has been questioned by Wierzbicka (1992), who claimed that it was merely a cultural artifact of English. For instance, Couper-Kuhlen (2009) investigated how disappointment was expressed through certain prosodic features, based on both British and German telephone conversations. She found that the vocal carriers for expressing disappointment included *ah*, *alright*, *oh*, *oh dear*, *oh I see*, *oh not to worry*, *oh right*, *oh well*, *oh well never mind*, *okay*, *okay never mind*, and *okay then*. Also, these vocal carriers were produced at a lower volume and with weaker articulatory force than usual, and with a low and slightly falling pitch. Couper-Kuhlen (2009) claimed that interpreting emotion was context-dependent. It was based on lexical and prosodic cues and judged with respect to a set of affects that were considered to be *relevant* for that specific location in the context. Furthermore, she also claimed that it was very likely that a specific linguistic device such as *oh* could sometimes communicate "disappointment" and at other times express "surprise", "annoyance", or "hesitation", depending a lot on its associated prosodic and linguistic features and its communicative context. Although emotions and attitudes are not the focus of the present study, they will be addressed as relevant to the studied micro-feedback and its prosody in this thesis.

### 3.5.2 Research on prosody of vocal-verbal micro-feedback in particular

A few researchers have studied the prosody of micro-feedback and the like. For instance, Ward (1996) and Ward and Tsukahara (2000) believed that *backchannel* (similar to *micro-feedback*) was not produced at random but was dependent on factors such as introduction of new information, syntactic completion of a grammatical clause, and prosodic clues from the earlier utterance, and they investigated especially the latter. The purpose of their study was to give dialogue speech systems a responsive and reacting ability. Ward (1996) used a corpus of 17 Japanese dyadic conversations between university students who were seated in such a way that they did not have eye contact. Ward and Tsukahara (2000) extended their study by including both these Japanese conversation data and also eight English conversations between American speakers who were seated in the same way without eye contact. They found that *backchannel* was not produced whenever the listener felt like it but was often cued, encouraged, or allowed by the speaker, for instance, by means of low pitch and rising intonation (see Ward & Tsukahara, 2000, p. 1203). Furthermore, Ward and Tsukahara suggested that 110 millisecond regions of low pitch were fairly good predictors of subsequent *backchannel*, and they also found that more obvious factors such as utterance end, rising intonation, and specific lexical items accounted for less than they intuitively seemed to (see Ward & Tsukahara, 2000, p. 1203). Although Ward (1996) and Ward and Tsukahara (2000) did not control for the gender, native dialect, education level, conversation location, mutual familiarity, and so on, which certainly play a role in the account of micro-feedback behaviours, their studies were regarded as big steps towards using prosodic clues to decide when to produce *backchannel* in an automatic dialogue system. Although Ward (1996) and Ward and Tsukahara (2000) did not investigate the prosodic features of micro-feedback itself, they suggested that micro-feedback and the prosody of relevance were correlated.

Tronnier and Allwood (2004) studied the duration and pitch contour of vocal-verbal micro-feedback, based on 40 excerpts of Swedish naturalistic dialogues in the Göteborg Spoken Language Corpus. They focused solely on the vocal-verbal micro-feedback *ja* (*yes*) and *nej* (*no*) in connection with agreement and disagreement, respectively. Tronnier and Allwood (2004) found that both falling and rising pitch contours were favoured for agreement, whereas more variable pitch contours were assigned to disagreement. Also, they found that the length of the vocal-verbal micro-feedback words varied between 81 and 462 milliseconds and that the long duration of a vocal-verbal micro-feedback word might strengthen its communicative function, for example, the emotional and attitudinal reactions to the communicated information (i.e., E/A in CPUE/A). In this thesis, the communicative function E/A of

micro-feedback is not the focus of the study but the function U is, that is, understanding in CPUE/A.

Cerrato (2005) conducted a study of the acoustic and prosodic characteristics of *m*-like sounds in Swedish, which are mostly micro-feedback expressions. She investigated the relationship between prosodic variation and communicative function of *m*-like sounds. The main hypothesis that different communicative functions of *m*-like sounds are conveyed by means of different prosodic cues was tested in two audio-recorded map task conversations and two video-recorded travel agency dialogues. Among other findings, Cerrato confirmed that the *m*-like micro-feedback expressions were used as acknowledgements<sup>29</sup> of comprehension (see Clark & Schaefer, 1989). This lends added support to the fact that there is some relation between micro-feedback and understanding, which will be further investigated in this study. Also, Cerrato found that *m*-like sounds mostly had a flat pitch contour when they were associated with hesitation and produced with disfluency, and that they had a falling and then rising pitch contour when they were expressing surprise. Cerrato's study provided a simple and sufficient coding scheme of pitch contour, which mainly includes rising, flat, and falling, and proposed that prototypical pitch contour and acoustic characteristics of micro-feedback were correlated to its communicative functions. This pitch contour coding scheme is used in the present study (see Section 4.1.9), and the prosodic features of micro-feedback are studied in relation to one of its communicative functions in particular, that is, understanding.

Romero-Trillo (2010) argued that the *pragmatic markers*, such as *yeah*, *okay*, *yes*, and *mhm* which are called *micro-feedback* in this study, were the explicit formulations of understanding. In his study, Romero-Trillo described how speakers indicate understanding through *pragmatic markers*, and also how *pragmatic markers* can realise other functions (e.g., emotions) with the exception of understanding. The analysis was based on a collection of spoken language data from interviews with intermediate to advanced speakers of English. The interviewees were always native English speakers. All the participants including both interviewers and interviewees were females of a similar age and similar educational background (university students). Romero-Trillo focused mainly on the initial pitch and the final pitch (computed by Praat) of the *pragmatic markers* that occurred in the interview conversations. He found that there was a difference in pitch level in the pronunciation of pragmatic

---

<sup>29</sup> According to Clark and Schaefer (1989), *acknowledgement* is method used by interlocutors to signal that a contribution has been understood well enough to allow the conversation to proceed (see the discussion also in Cerrato (2005, p. 21).

markers between the native and non-native speakers of English in that the non-native speakers' lower level of especially the final pitch, in comparison with the natives', indicated more of agreement or believing than just giving a response or feedback. Romero-Trillo's study not only strengthens the relationship between micro-feedback and its communicative functions of understanding and emotional and attitudinal reactions (i.e., UE/A), but also suggests that there is some correlation between the prosody of micro-feedback (although pitch only in Romero-Trillo's case) and its UE/A functions. Of these functions, U (i.e., understanding) is the focus of this thesis.

### **3.6 Focus of the present study**

As stated earlier, little has been done to investigate how understanding can be identified, classified, operationalised (by using both qualitative and quantitative methods), and analysed in empirical interaction data. Likewise, little has been done to research the relation between micro-feedback and understanding, for example, how (much) micro-feedback can function as signals for understanding, that is, in which way and to what extent which specific type of micro-feedback conveys what kind of understanding. There is a need to study these in greater detail. What understanding difficulties and problems are and how they are coped with in intercultural communication with empirical data from both FTF and VMC interactions will be investigated.

In this thesis, understanding and understanding problems are investigated by focusing on micro-feedback and other related responsive communication behaviours, primarily, meaning repair, as regards how understanding and understanding problems are coped with in communication. Understanding will be conceptualised, with respect to how it can be subdivided, how it occurs, how it is detected, handled, and resolved, how it constructs sense-making and information sharing, and what can be compared between different communication contexts. The features and patterns of understanding and understanding problems will be explored in the empirical data in primary Study 2. More details of the annotation and coding scheme of understanding are presented in Section 4.2.7.

Micro-feedback will be studied in terms of its modality (i.e., auditory and visual modality) and prosody (mainly pitch and duration) features in relation to sufficient understanding, misunderstanding, and non-understanding in Swedish–Chinese intercultural communications. Micro-feedback, involving auditory and visual modalities, includes the unimodal vocal-verbal, the unimodal gestural, and the multimodal micro-feedback. Pitch mainly includes the maximum, minimum, and mean pitch values (in Hz), the pitch range values and types (i.e., small, medium, and large), and

the pitch contours (i.e., falling, flat, and rising). Duration includes the duration values (in milliseconds) and duration types (i.e., short, medium, and long). The modality and prosody of micro-feedback will be explored in relation to understanding in primarily Study 1. More details of the coding schemes and annotations are presented in Section 4.1.9.

As a consequence of these classifications and measurements, the operationalisation of understanding and communicative micro-feedback can be further enriched. As a result, better insights into the features of understanding and micro-feedback and the relations between them can be described.



# 4

## Method

Study 1 and Study 2 have been conducted in order to address the research questions. In Chapter 4, the methodological issues concerning data collection and analysis are described.

### 4.1 Method of Study 1

Data has been collected from eight face-to-face strangers' conversations between four Swedish and four Chinese speakers. The data has been studied in two empirical analyses. The first analysis focuses on the modality of micro-feedback including both unimodal (i.e., vocal-verbal or gestural) and multimodal (i.e., vocal-verbal plus gestural) expressions. The second analysis investigates the prosodic features (i.e., pitch, pitch range, pitch contour, and duration) of the vocal-verbal micro-feedback. Many other vocal-verbal and gestural interactions occur in the data, but only those that make up the micro-feedback items are examined in this study.

#### 4.1.1 Intercultural communication between Swedish and Chinese

As introduced earlier in the literature review, culture differences make it more difficult to approach mutual understanding in communication (see e.g., Gogan et al., 2014; Gumperz, 1982; Pride, 1985; Tannen, 1990; Smith, 2004; Eliot, 2010; Samovar et al., 2012). On the one hand, the Swedish culture from Western Europe and the Chinese culture from Eastern Asia are considered very different in many physical,

regional, linguistic, and social aspects. Thus, it is assumed that understanding problems or different types of understandings, for instance, misunderstanding, and non-understanding, may occur with a high frequency in such an intercultural communication situation. On the other hand, the world is a global one, and Sweden and China have more and more international cooperation and communication. It is therefore important and useful for people to get to know more about how Swedes and Chinese understand one another.

#### 4.1.2 Spontaneous FTF dyadic communication

When people communicate, they use a variety of means, such as telephone, email, video conferencing, and so on, involving both spoken and written forms and varying in terms of being spontaneous or not. Spontaneous communication has a strong impact on mutual understanding and future cooperative activity. It is unplanned, informal, and not necessarily related to any specific issue, and it is very much dependent on FTF *co-presence* (Rapp & Jackson, 2003, p. 236).

Equally important, FTF communication is historically seen as the basis of a theory of language, the basis of all human language behaviour, and the standard communication situation (Clark, 1996). As Hutchby (2001) suggested, FTF communication is historically first, most common, most widely spread, and evolutionarily adapted. FTF communication involves more than just words. It involves many sensory modalities (e.g., auditory and visual) and minimises the efforts required to communicate, and it is also a joint action between communicators in a more easily *cooperative* way (Monk, 2009, pp. 13–14). In one fundamental sense, it takes place in one common communication situation that provides more opportunities for *feedback* (McHale, 2004, p. 206). Spontaneous FTF communication serves the purpose, for this thesis, of studying understanding and its communicative signals of micro-feedback in communication.

Also, in dyadic communication situations, two communicators are more likely obliged to or engaged in stimulating and communicating with each other (Ruffner & Burgoon, 1981) and tracking communicative gestural behaviours such as head nod or eyebrow rise may be more difficult with more participants (Anderson, 2006). Thus, dyadic communication is seen as maximising the opportunities for interlocutors' self-disclosures in their most natural habitat in terms of their communicative behaviours. In this thesis, eight Swedish–Chinese spontaneous FTF dyadic dialogues are studied.

### 4.1.3 Between unacquainted university students

The participants are four Swedish and four Chinese university students (with an equal number of male and female students of each cultural group), who were strangers to each other and were studying in Sweden. They were on average 26 years old, and they had normal human communication abilities. Thus, the participants are comparable in the sense that they had a shared social background of being university students.

Also, getting acquainted is one typical interpersonal communication in social interaction (Svennevig, 1999). Svennevig (1999) has emphasised that the first encounter is a recognisable communication activity with specific procedures involving certain resources and constraints in terms of how interlocutors contribute to establishing a certain interpersonal relationship. He has pointed out that how interlocutors introduce and develop their first conversation is often done by eliciting questions such as *where do you come from* or *what do you do* that introduce a conversational sequence and involve certain constraints on the expected responses and the subsequent development of the conversation. According to Svennevig (1999), unacquainted people engage in interrogative behaviour (i.e., asking questions to acquire information about the “other”) and self-disclosure (i.e., providing information about the “self”), thus enabling them to proceed and coordinate in the conversation and enhance their mutual understanding.

Also, Maynard and Zimmerman’s (1984) study of the acquainted and the unacquainted interlocutors’ conversations showed that with a focus on establishing some kind of social and personal relations and achieving some common communication goals, the unacquainted interlocutors especially used a considerable number of questions and answers to initiate and develop topics and understandings that involved a lot of contextualised micro-feedback. That is, unacquainted people have a greater need for mutual understanding and more opportunities to elicit and give micro-feedback. This serves the research purpose of Study 1 to unfold the relationship between understanding and micro-feedback. Therefore, strangers who had no prior mutual acquaintance were selected and paired in each dialogue for Study 1 in this thesis.

### 4.1.4 A simple communication task

The four Swedish and four Chinese participants were instructed to communicate with a simple task of getting to know one another, preferably within eight minutes. They were given freedom as regards, for example, what they wanted to talk about and

what they wanted to know about. How they interacted and got to know and understand one another in this specific communication situation is investigated in the present study.

#### 4.1.5 Communication language and the participant's language competence

As the only common language between the Swedish and the Chinese participants, English was selected as the communication language.

The English language in this thesis does not refer to any British English, Australian English, American English, or Canadian English, but to English as an international lingua franca, which is widely used in an enormous variety of social and cultural contexts (see Mauranen & Ranta, 2009). As a communication tool, the English language (lingua franca) makes it possible for people with different cultural and language backgrounds to achieve mutual understanding and co-construction of communication.

The Swedish and the Chinese participants studied in this thesis have relatively good English language skills. First, in order to be eligible for university studies in Sweden, a student (both native and international) must meet the English language entry requirement. For instance, a Swede must have successfully completed certain upper secondary school (high school) studies or certain university studies in English (e.g., English 6 or English Course B) and a non-Swede must have an internationally approved English test (e.g., IELTS or TOEFL). This means the participants have at least theoretically good English skills. Second, the participants had been studying for a while at the Swedish universities, at least more than 6 months, before they participated in this research project. Since English is the instruction language for most courses and study programs in the universities in Sweden, the participants must have been good enough in practice to communicate and understand one another in the English language.

#### 4.1.6 Audio- and video-recorded dialogue data

The conversations were both audio- and video-recorded. In order to eliminate as much as possible the influence of factors such as physical environment and artifacts, the participants were recorded in a studio and in a standing position (see Figure 4.1). In this way, the external influences are minimised and the gestural micro-feedback is captured to a maximum extent. Besides, it is not much different from an ordinary strangers' meeting situation, since it is not uncommon that two strangers meet and start communicating in a standing position.



*Figure 4.1.* One Chinese participant in two conversations (filmed from the centre-positioned camera).

Three cameras (left, centre, and right-positioned) were used for the recording. Portable microphones were not used, because whether with or without a portable microphone, the participants in the dialogue move and gesture a lot all the time which affects the prosodic sound quality (e.g., intensity). Also, the main motivation for the project with this setup is primarily for studies of spontaneous FTF interactions, thus capturing vocal-verbal and gestural behaviours in such a naturalistic communication situation is more important. Nevertheless, apart from intensity, the main prosodic features such as pitch and duration are of good quality.

*Table 4.1.* The size of the analysed data in Study 1.

<b>Dialogue No.</b>	<b>No. of words</b>	<b>Time length (min.: sec.)</b>
1	2,070	11:44
2	1,380	07:56
3	1,309	09:04
4	1,555	10:29
5	1,070	06:52
6	1,122	08:11
7	943	06:08
8	678	04:44
<b>Total</b>	<b>10,127</b>	<b>65:08</b>

The video-recordings last between 4:44 and 11:44 minutes (with a mean of 8:08), consisting of 678 to 2,070 vocal words (with a mean of 1,266). In all, the recordings last 65:08 minutes and consists of 10,127 vocal words (see Table 4.1). In this thesis, *word*<sup>30</sup> instead of morpheme or phoneme is the smallest speech segment annotated and analysed. Each entire dialogue is analysed in the present study.

#### 4.1.7 Acoustically processed prosodic data

In order to study the prosodic features of micro-feedback, 661 instances of vocal-verbal micro-feedback, which did not overlap other sounds but were of good acoustic quality, were selected out of the 1,288 micro-feedback expressions in the entire interaction data. Among the 661 vocal-verbal micro-feedback expressions, there are 634 cases, 9 cases, and 18 cases with sufficient understanding, misunderstanding, and non-understanding, respectively.

The segmentation of the analysed prosodic data was performed with the help of the digital audio editor WaveLab<sup>31</sup> (Steinberg, 2002). Since most of these analysed vocal-verbal micro-feedback expressions occur at the beginning of each *contribution*<sup>32</sup>, the segmentation work is relatively straightforward. Normally, the onset is set at the point of the appearance of energy, while the offset is marked at the point of the disappearance of energy. However, because the energy changes are always dependent on the environmental sound interference (if any) and the participants' bodily movements (which affect the participants' physical distances from the recording microphone, the speech sound waves, and the sound intensity), the segmentation work also depended a lot on the annotator's auditory perception. Besides this, all the analysed prosodic clips were generated by WaveLab one dialogue at a time and named automatically as, for example, dial 1 (001).wav and dial 1 (001).gpk. Only the "wav." files

---

<sup>30</sup> In English, one *word* is operationalised as a sequence of graphs between two spaces occurring in the transcribed utterances of the interaction (see Lu & Allwood, 2011). That is, words are defined with regard to conventional writing.

<sup>31</sup> *WaveLab* is a digital audio editor by Steinberg aimed at the professional as well as the semi-professional market. It supports multi-channel files, DirectX plugin, VST plugin and DVD-Audio creation. WaveLab was started in 1995 mainly by programmer Philippe Goultier. Cut-down versions of WaveLab include WaveLab Studio, WaveLab Essential and WaveLab LE.

<sup>32</sup> Based on GTS (Nivre et al., 2004) and MUMIN (Allwood et al., 2007), which are used in this thesis, *contributions* are the transcribed utterances and communicative body movements (i.e., gestures of the participants) in the recorded communication activity. A *contribution* always belongs to a single participant. This means that even if two participants utter a phrase collectively (overlapping simultaneously or by uttering one part each), this will count as two separate (individual) contributions (see GTS). Each contribution has an interactive relation (i.e., *responsive-initiatory relation*); that is, each contribution responds to something prior and gives rise to some possible next (see Linell, 2009, p. 212).

were imported into Praat afterwards. In the meantime, one “MRK” file, which contains the markings of all the prosodic clips, was generated automatically by WaveLab for each dialogue. Each “MRK” file was kept, in order to easily trace back where in particular a specific sound clip was extracted from the specific and entire dialogue.

The measurements of duration and pitch features such as the duration value and the pitch of each sound clip in the prosodic data were processed and generated automatically by Praat<sup>33</sup> (Boersma & Weenik, 2009). The relevant values were measured from both spectrograms and waveforms. The annotation of the pitch contour patterns was performed manually by independent annotators (see Sections 4.1.9 and 4.1.10). Given that Praat can automatically produce contour drawings, the pitch contour shape of each micro-feedback sound clip can easily be obtained. However, this visualised contour shape by Praat may not reflect the final annotation of the contour studied. For instance, there are cases identified by Praat with clearly sharp rising and falling pitch contour visualisations and there are many cases that do not have sharp or distinct Praat contour shapes (see examples in Section 4.1.9: flat *yeah* of dial 2 (033).wav. in Figure 4.4, falling *okay* of dial 1 (060).wav. in Figure 4.5, falling *aha* dial 1 (035).wav. in Figure 4.6, and rising *eh what do you mean* of dial 7 (040).wav., falling *okay* of dial 1 (049).wav., and falling *of course* of dial 1 (066).wav. in Figure 4.7). In these cases, it becomes difficult to annotate the pitch contour pattern by just interpreting the Praat automatic visualisations. Therefore, in this thesis, over and above the Praat visualisation of contour shapes, the annotation of the pitch contour was primarily carried out according to how it was perceived acoustically by the independent annotators (see more in Section 4.1.9). The visualisation from Praat and the auditory perception from independent coders can provide better insights of the prosody data. Study reliability can be increased. Duration was measured in milliseconds, and was categorised into three duration types (see Section 4.1.9). The pitch values, which include maximum pitch, minimum pitch, and mean pitch, were measured in Hz. The pitch ranges were classified accordingly (see also Section 4.1.9). The value distribution procedure is not perceptual (e.g., based on the duration or pitch range perception) but mechanical (see more details in Section 4.1.9).

---

<sup>33</sup> *Praat*, which means “talk” in Dutch, is a freeware program for the analysis and reconstruction of acoustic speech signals, which is widely used in phonetics. It was developed by Paul Boersma and David Weenink at the Phonetic Sciences department at the University of Amsterdam (see more details in Boersma and Weenink, 2009). It can be downloaded from the website: <http://www.fon.hum.uva.nl/praat/>. Praat is a very flexible tool for carrying out speech analysis. It offers a wide range of standard and non-standard procedures, including spectrographic analysis, articulatory synthesis, and neural networks.

A number of issues concerning prosody analysis should be mentioned here. Prosody is individual, gender, and context dependent. It is not uncommon that people tend to converge or adapt their prosody when they communicate with one another. People have different basic pitch levels (e.g., man and woman, old and young, superior and subordinate). Also, individuals have various rates of speech. Some speak faster and others speak slower. Equally important, Chinese and Swedish are two languages that use prosody phonemically and the participants were speaking a third language, the English lingua franca, in which prosody does not play the same role. Also, individuals in even the same culture have different phonemic characteristics. These questions should be taken into account when conducting a prosody analysis.

In the present study, however, because prosodic features are not analysed or compared with regard to different individuals or groups, these issues and the related empirical results were not normalised as they normally are in phonetic studies. Besides, there are equal numbers of male and female participants in the study and the prosodic features of all participants in common are of research interest, and the normalisation problem is minimised. Without normalising the prosody data, the current study method increases the variety of the data and probably also decreases the chance of finding any difference. However, it does not bias the data or invalidate the result.

#### 4.1.8 Transcription standard

The recorded data was manually transcribed and checked according to The Göteborg Transcription Standard (henceforth GTS) version 6.4 (Nivre et al., 2004) created at the Department of Linguistics of the University of Gothenburg in Sweden. GTS is primarily used as a standard for machine-readable transcriptions of spoken language. It was first used within the research program of Semantics and Spoken Language at Göteborg University. In the last decade, it has been more popularly used as a transcription standard for the study of spoken language features and social activity patterns.

GTS was used in this present study, for two primary reasons. First, the first empirical work in this thesis (Study 1) was included in the NOMCO corpora<sup>34</sup>. For the purposes of further developing the transcribing and coding systems for multimodal communication studies and making comparative analysis, the same standards were

---

<sup>34</sup> The NOMCO project refers to the Multimodal Corpus Analysis in the Nordic Countries in the 2000s at the universities of Gothenburg, Copenhagen and Helsinki. The focus was to study communicative phenomena such as feedback, turn management and sequencing in different social interactions such as first encounters, group meetings, and job interviews. Study 1 in this thesis was included in the Nordic multimodal first encounters corpora.

employed for the project's related studies. Second, GTS supports automatic machine learning, which makes quantitative counting and frequency comparison possible and efficient for linguistic analysis. Besides, applying machine learning techniques to create support for automatic recognition of gestures with different communication functions was one of the research aims at that time.

The terms *transcription*, *annotation*, and *coding* are used from a linguistic perspective in this thesis. Based on the GTS (Nivre et al., 2004), the MUMIN (Allwood et al., 2007), and the INTSINT (see Hirst & Di Cristo, 1998; Hirst, 1999), these three terms can be explained as follows. Traditionally, in linguistics, *transcription* refers to the representation of speech in written form. In the present study, for reasons of language economy, transcription is used to refer to both the written form of the speech utterance and the written description and annotation of the relevant gestural communicative behaviours and contributions. *Annotation*, in linguistics, includes comments and metadata, which are interpreted by annotators. A collection of texts with linguistic annotations is known as a corpus (plural corpora). Tools and formats are normally needed to create and manage linguistic annotations. In this thesis, the functions of micro-feedback such as C (contact), P (perception), U standing for sufficient understanding, -U for non-understanding, and misU for misunderstanding, as well as E/A (emotional and attitudinal reactions, for example, surprise, amusement, and hesitation) and evaluative opinions such as agreement and disagreement are annotated. Furthermore, the three types of pitch contours (i.e., rising, flat, and falling) are annotated, according to both the Praat contour visualisation and how the acoustic sound is perceived by the independent annotators. *Code* is simply a rule for converting a piece of information (e.g., a letter, word, phrase, or gesture) into another form or representation like one sign into another that is not necessarily of the same type. Accordingly, *coding* refers to the analytical process in which data are categorised to facilitate the analysis, in both quantitative (such as frequency figures and questionnaires' statistical results) and qualitative forms (such as interaction transcriptions and interview transcripts). Therefore, in the present study, micro-feedback is coded in terms of modality, and VFB and GFB are codes of vocal-verbal micro-feedback and gestural micro-feedback, respectively. Duration type (i.e., short, medium, and long) and pitch range type (i.e., small, medium, and large) are also coded.

In the GTS system, there are standards for how to transcribe pause and silence in the dialogue. For example, /, //, ///, the number of slashes indicate the length of a pause, corresponding to short, intermediate and long pauses, respectively (see more descriptions in Appendix A). A vertical bar | indicates silence. Note that silences are not pauses but simply the moments when time passes without anybody saying or doing anything. The symbol < | > indicates a pause where communicative gestures

are inserted or occur in the interaction. Pauses and silences in communication are not focused on in this thesis, however, they are transcribed and annotated much more thoroughly in Study 1 than in Study 2. When a pause or silence is interpreted as associated with certain responsive reactions (e.g., emotional and attitudinal responses of hesitation and uncertainty), it is taken into account in the present thesis<sup>35</sup>.

#### 4.1.9 Annotation and coding schemes

In this thesis, a number of annotation and coding schemes are used for micro-feedback, understanding, pitch contour, duration, and pitch range.

##### Micro-feedback annotation

A variant of the MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena (abbreviated as MUMIN) (Allwood et al., 2007) was used in this thesis. MUMIN was originally created to experiment with annotation of multimodal communication in video clips. Over the last ten years, it has been popularly used as a general instrumental framework, especially among Nordic researchers, for the study of gestures and facial displays in interpersonal communication, in particular with a focus on the multimodal expressions for *feedback*. Based on some earlier studies that have been conducted (e.g., Lu, 2010; Lu & Allwood, 2011; Lu, 2012), a few more gestural types than those already included in MUMIN were identified and further noted. Thus, a variant of MUMIN was adopted in this thesis.

---

<sup>35</sup> The results of the present thesis suggest that a pure pause or silence without any hesitating or uncertain vocal sound or gestural behaviour and without any emotional and attitudinal responsive action barely occurs in the data analysed. Thus, it is hardly associated to understanding in this study.

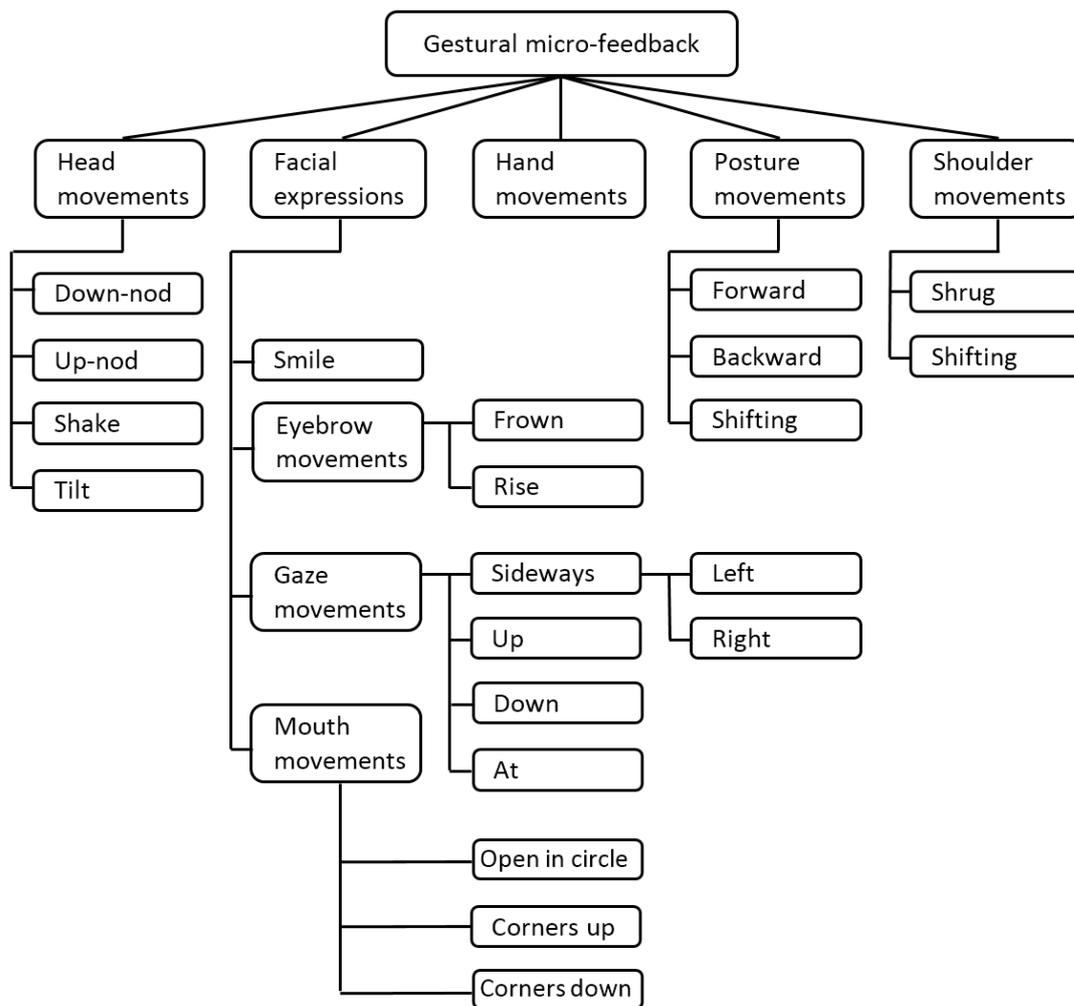


Figure 4.2. Classification of gestural micro-feedback.

The gestural micro-feedback consists of head movements, facial expressions, hand movements, posture movements, and shoulder movements (see Figure 4.2). Among them, micro-feedback head movements include head *nod*<sup>36</sup> (i.e., *down-nod*), *up-nod*, shake, and tilt; micro-feedback facial expressions include smile, eyebrow movements (e.g., frown, rise), gaze movements (e.g., gaze sideways (i.e., gaze to the left or right), gaze up, gaze down, gaze *at*<sup>37</sup>), and mouth movements (e.g., open in a circle, corners

<sup>36</sup> In this thesis, *nod* is used particularly as a synonym of *down-nod*, referring to a head movement that starts from the neutral position and goes downward. Similarly, *up-nod* denotes a movement upward from the neutral position. When repeated instances of these expressions occur in quick succession, the plural form *nods* is used.

<sup>37</sup> *Gaze at* refers to gazing towards the other interlocutor. In this thesis, non-communicative or meaningless gazing at or other gaze movements, mouth movements, and blinks are not regarded as anything more than natural human actions, which one person always does when he or she is awake.

up, corners down, and others). Due to the complexity of micro-feedback hand movements, they are not subcategorised in depth in this study. Rather, hand movement is identified when it is related to micro-feedback and coded in a descriptive way based on how it actually takes place in the communication activity. The micro-feedback posture movements primarily refer to the *marked postures*, that is, trunk movements and posture changes (e.g., posture forward, backward, and shifting). The micro-feedback shoulder movements include, for example, shrug and shoulder shifting. In this thesis, single and repeated gestures are coded separately (e.g., nod and nods, up-nod and up-nods, shrug and shrugs). Because the gestural micro-feedback occurs very often together with the vocal-verbal micro-feedback (see Lu & Allwood, 2011), the vocal-verbal micro-feedback is a helper to distinguish micro-feedback gestures from other gestures.

Equally important, the vocal-verbal micro-feedback primarily consists of micro-feedback words, such as *mhm, aha, yeah, okay*, micro-feedback phrases such as *yeah yeah, no no no no, all right*, and micro-feedback sentences such as *Yeah yeah yeah you're right, Really? I cannot believe it, Sorry pardon? What did you say?, Yes, I agree with you*. Different types of micro-feedback are examined in this study.

## Understanding annotation

As discussed earlier, overt and manifested understanding is what will be studied, and it will be interpreted primarily from the analyst's perspective in this thesis. The so-called fake or pretended understandings, which are not observed or detected by the annotators, are not taken into account.

Primarily based on Allwood's (1986), Weigand's (1999), and Zaefferer's (1977) ontological theories and frameworks and Garfinkel's (1967), Bakhtin's (1986), Linell's (2009), and Lindwall and Lymer's (2011) suggestions concerning classifying understanding, a set of criteria of understanding, which includes sufficient understanding, misunderstanding, and non-understanding, is used in the present study.

As discussed already in Chapter 2 of this thesis, *sufficient understanding* occurs when the understanding is sufficient to serve current practical purposes (Garfinkel, 1967; Bakhtin, 1986; Taylor, 1992; see Linell, 2009, p. 222) of information sharing, sense-making, and continuing communication, regardless of whether the understanding is full or partial (see Linell, 2009). In sufficient understanding, interlocutors are content with understanding one another and it is good enough to proceed further in the communication (see Lindwall & Lymer, 2011). *Misunderstanding*, in this thesis, is defined as an insufficient understanding, in that although it can perhaps serve current communication purposes, the information is understood in an incorrect way

deviating from what is intended or anticipated (Weigand, 1999). That is, misunderstanding is unintended or unanticipated. The interlocutor who misunderstands is not aware of it (Weigand, 1999). *Non-understanding* is the opposite of sufficient understanding, in that it cannot serve the current communication purposes regarding sharing and making achieved sense of the information that is presented. Non-understanding can be a static or a gradual and dynamic process, which means that there can be a total non-understanding and a lack of understanding, which leads to a situation in which more or less nothing is understood.

Both non-understanding and misunderstanding are regarded as insufficient understandings (cf. Zaefferer, 1977) in the present study. However, non-understanding differs from misunderstanding in the sense that it does not involve any achieved sense of the communicated information. In contrast, misunderstanding does involve sense-making, although in an incorrect way in relation to what the communicator intends and anticipates. By using these criteria, the empirical data of understanding studied is annotated as sufficient understanding, misunderstanding, and non-understanding, and they are coded as U, misU, and -U, respectively. Various degrees of each of them are not taken into account in this thesis. Detection of understanding problems is approached from the participant's perspective. Annotation of understanding problems is done from the analyst's perspective.

### Pitch contour annotation

With regard to annotating the pitch contour of vocal-verbal micro-feedback, Tronnier and Allwood (2004) made an attempt to relate it to the pitch contour of not only the previous vocal word but also the following one. For example, Tronnier and Allwood's coding scheme for pitch contour includes falling  $F_0$ , steady  $F_0$  followed by a rise, rising  $F_0$ , oscillating  $F_0$ , and fall and rise.

Cerrato (2005) had a simpler coding scheme for pitch contour, which primarily includes rising, flat, and falling, and also combinations of them such as falling-rising. In the present study, the annotation of the pitch contour focuses on comparing the last syllable of the vocal-verbal micro-feedback expression with the others in the duration of the micro-feedback. Thus, various pitch contour combinations such as falling-rising, flat-rising, and the like are not taken into account in this thesis.

A simple set of criteria to code the pitch contour of vocal-verbal micro-feedback, which includes rising, flat, and falling, is primarily used instead (see Figures 4.3 to 4.5). Illustrations of rising pitch contours are presented in Figure 4.3, and those of flat and falling pitch contours are shown in Figures 4.4 and 4.5, respectively.

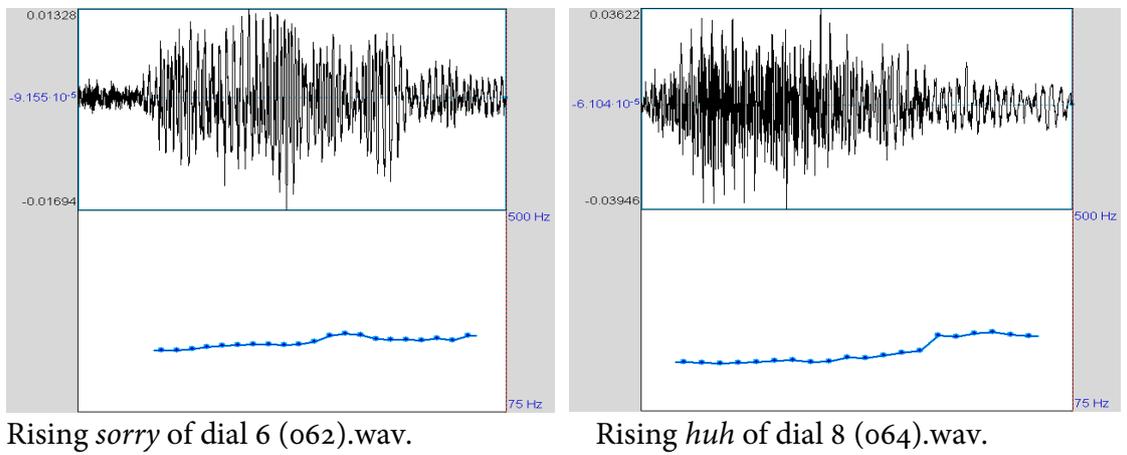
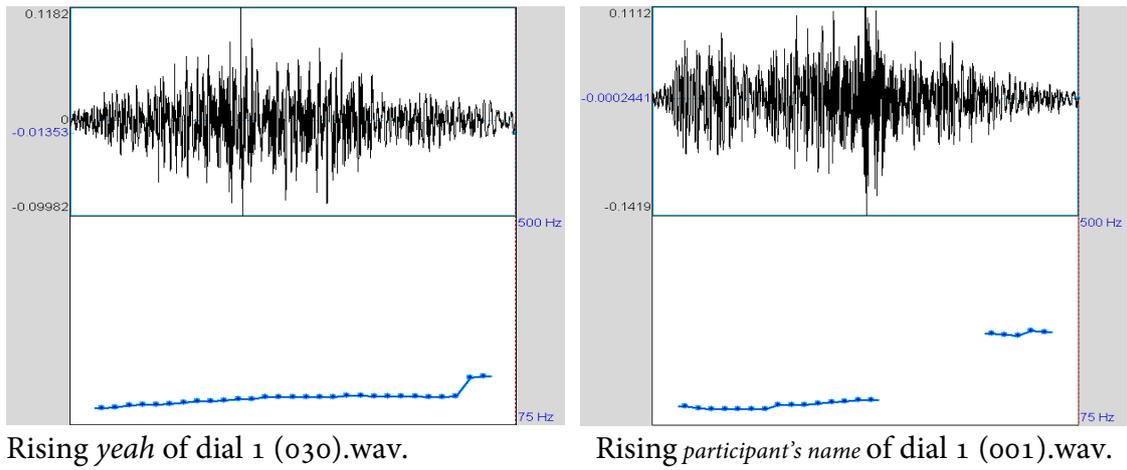


Figure 4.3. Illustrations of annotated rising pitch contours.

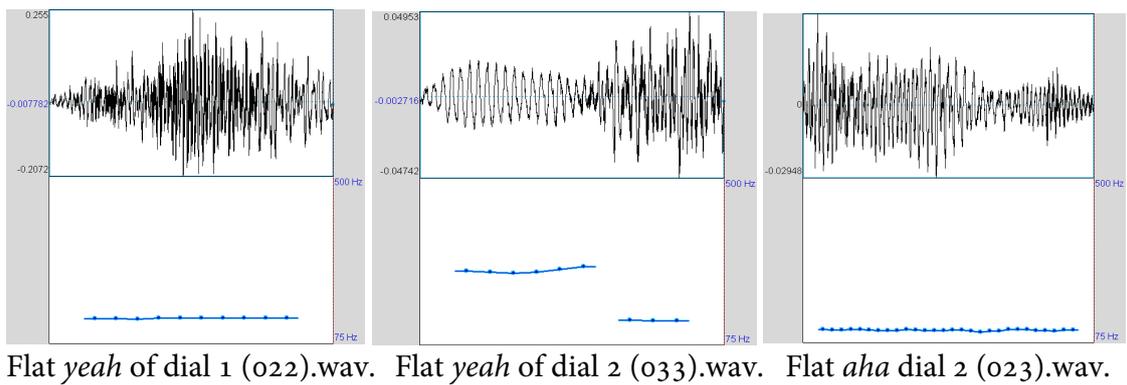
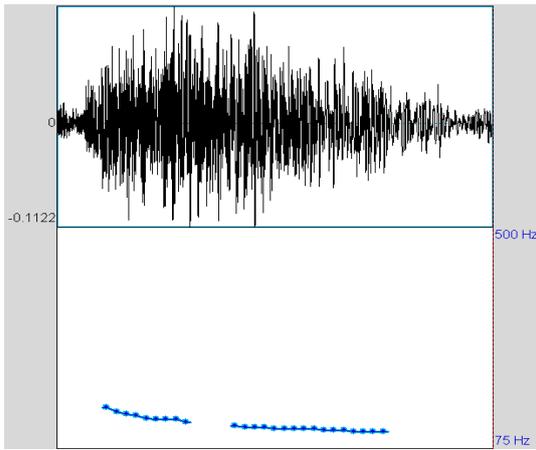
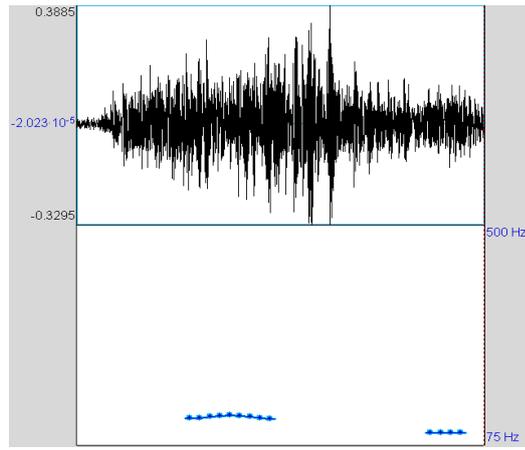


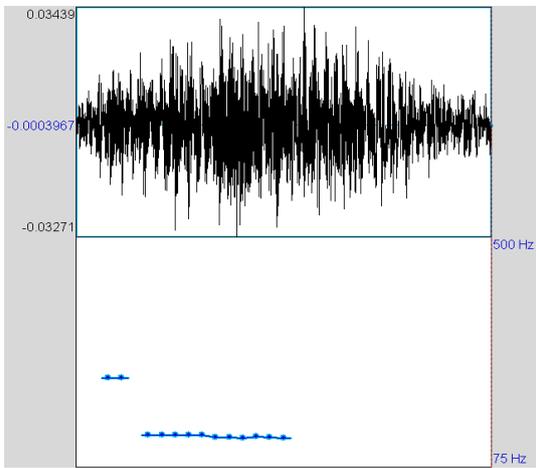
Figure 4.4. Examples of *yeah* and *aha* that are annotated with a flat pitch contour.



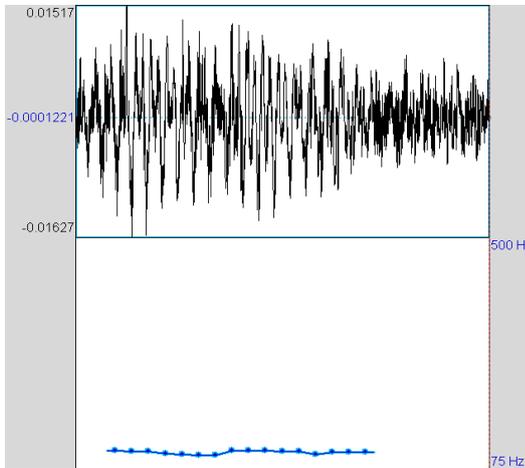
Falling *yeah* of dial 1 (100).wav.



Falling *yeah yeah yeah* of dial 1 (026).wav.



Falling *ah* of dial 1 (029).wav.



Falling *okay* of dial 1 (060).wav.

Figure 4.5. Illustrations of the contours annotated as falling (examples of *yeah*, *yeah yeah yeah*, *ah*, and *okay*, respectively).

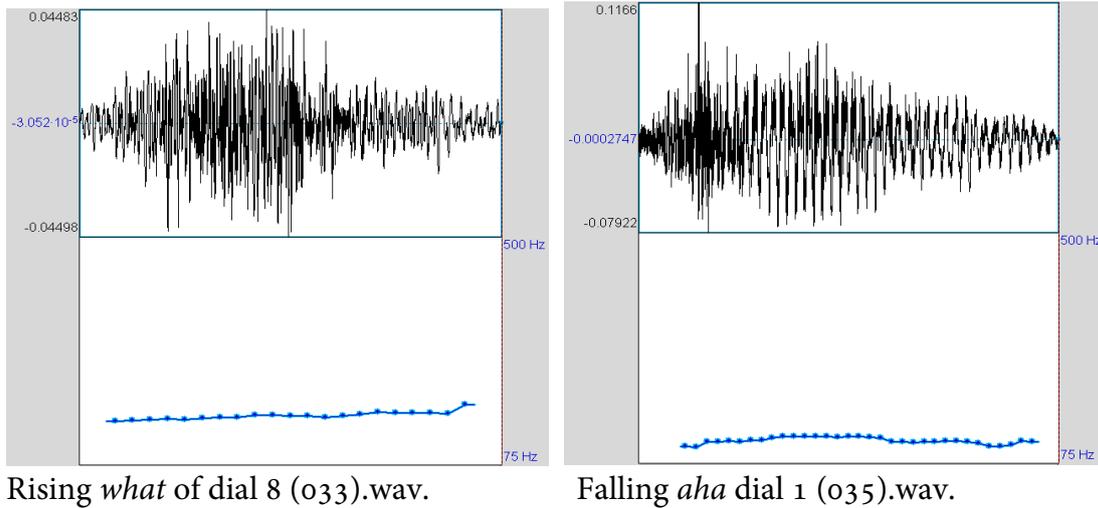
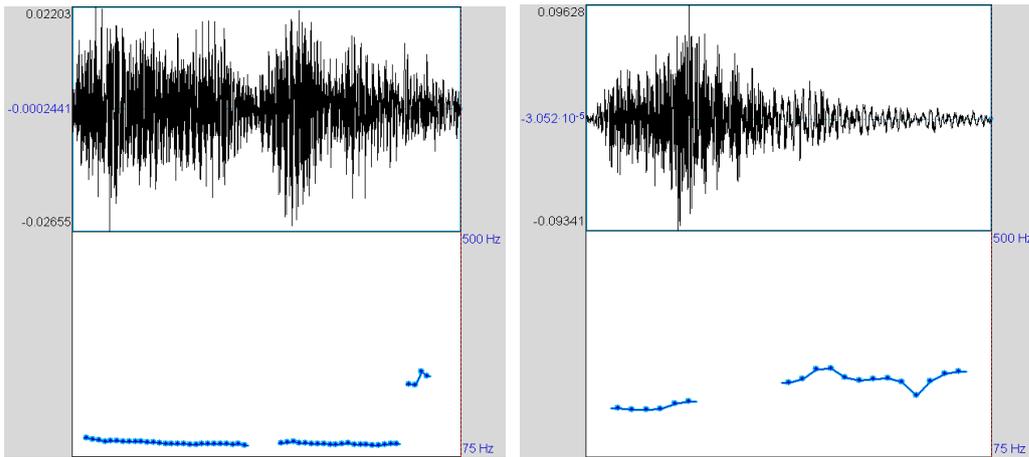


Figure 4.6. Annotated rising pitch contour and falling pitch contour with the contour visualisations of displaying some flattening.

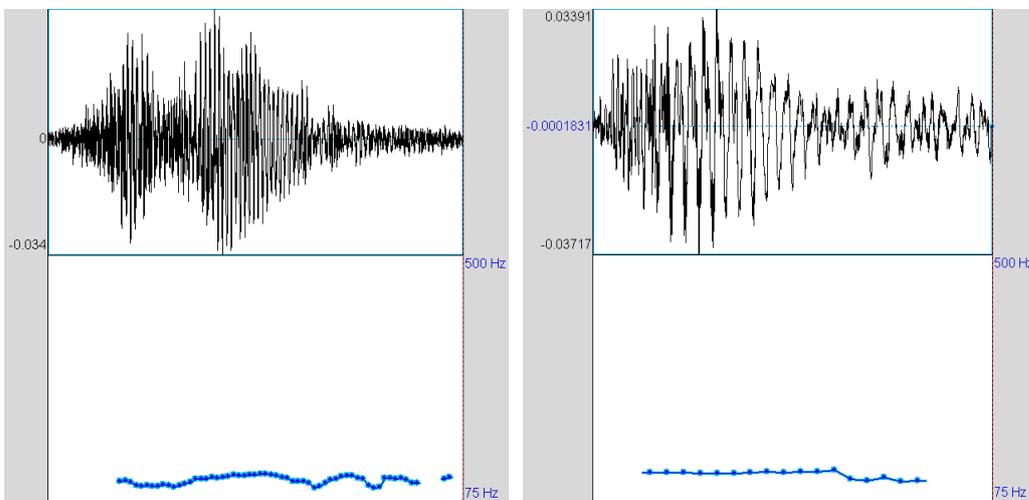
As mentioned above, the annotation was not only based on the contour visualisation which was automatically generated by the software Praat but was also very much dependent on how it was perceived acoustically by the independent annotators (independent annotation procedure is presented in Section 4.1.10). See more illustrations below.

Some of the contours seem to have a more or less flat tendency in the visualised contour shape, but they are acoustically perceived to be something else, for example, rising or falling (see Figure 4.6). In such cases, they are annotated according to the annotator's auditory perception as rising or falling rather than flat.

There are also some contours that occur in a dramatically changing shape in their contour visualisations produced by Praat (see Figure 4.7). Cases like these were annotated primarily according to what was acoustically perceived by the independent annotators rather than simply relying on Praat's contour drawings.

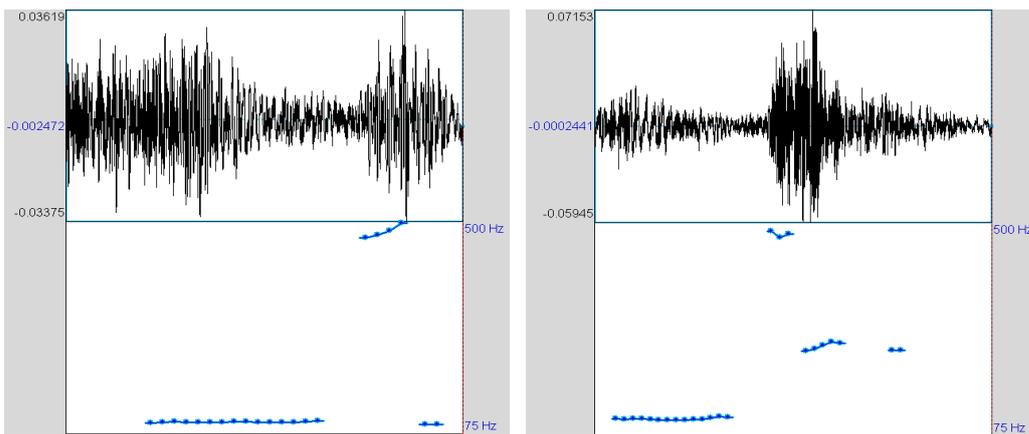


Rising *eh* what do you mean of dial 7 (o40).wav. Rising *huh* of dial 5 (o60).wav.



Flat *mhm* of dial 2 (o36).wav.

Flat *okay* of dial 2 (o46).wav.



Falling *okay* of dial 1 (o49).wav.

Falling *of course* of dial 1 (o66).wav.

Figure 4.7. Contours of dramatic changes in the visualisation shape, which are annotated according to the annotators' perceptions (*Note*. Rising, flat, and falling in the figure are the perceived and annotated contours).

### Duration coding: duration values and duration types

The duration of vocal-verbal micro-feedback is studied in terms of both its values and its duration types. For the former, the numerical values are produced automatically by Praat, and they are measured in milliseconds. For the latter, Hirst and Di Cristo (1998) and Hirst (1999) proposed a scalar feature of duration with five degrees: extra-short, short, normal, long and extra-long. Although each sound clip was segmented in the same way as in this thesis, the duration of the segment involved an

$$\hat{d}_{ru} = \left( \sum_{i=1}^m (\bar{d}_{ip}) + k * q \right) * t$$

algorithm like  $\hat{d}_{ru}$ , specifically designed for synthesis of speech prosody. Since it is more of complex (e.g., in terms of the mathematical principles) than necessary to serve the purpose of the present study, this duration type coding scheme is not used in this thesis. Xu and Wang (2009, p. 506) proposed a theory of *minimum, maximum, mean, and interval between minimum and maximum* for phonological measurement. It seems that Hirst and Di Cristo (1998), Hirst (1999), and Xu and Wang (2009) all have a tendency to focus on the two far ends and the in-between of the phonological item when making the classification. Thus, a simpler categorisation of duration type is made as follows.

All the duration values are divided into three categories, that is, short, medium, and long, with an even distribution. This distribution procedure is not perceptual (e.g., based on the duration perception) but mechanical. First, it is found that the minimum and maximum of the entire duration data are 82 and 1748 milliseconds, respectively. Then, all the possible values in-between the minimum and the maximum are evenly distributed into three groups. That is, 1748 minus 82 divided by 3, which equals 555. Following this, the three categories of duration types are based on the intervals between 82 and 637 (82+555) as short duration, between 638 and 1192 (637+555) as medium duration, and between 1193 and 1748 as long duration.

### Pitch range coding: pitch range values and types

Each vocal-verbal micro-feedback is also studied with respect to its pitch values (maximum, minimum, and mean) and the relevant pitch range values. Each pitch range value equals the relevant maximum pitch value minus the minimum pitch value.

Regarding the coding of the pitch range, the INTSINT<sup>38</sup> coding system (see Hirst & Di Cristo, 1998; Hirst, 1999) categorised it into three types: top, mid, and bottom.

---

<sup>38</sup> INTSINT is an acronym for INternational Transcription System for INTonation. It is a coding system for intonation, originally developed as a prosodic equivalent of the International Phonetic Alphabet. It was developed by Daniel Hirst and his colleagues at the CNRS Centre of the Aix-en-Provence University

Besides this, referring to Xu and Wang's (2009) theory for phonological measurement, the categorisation of the pitch range type used in this thesis includes small, medium, and large. The categorisation is performed as follows.

This distribution procedure is not perceptual (e.g., based on the pitch range perception) but mechanical. First, with the help of Praat, it is found that based on the entire prosodic data, the maximum pitch range value is 459 Hz and the minimum is 0 Hz. Then, between the minimum and the maximum pitch range values 0 and 459 Hz, three even groups with a division mean of  $459/3$  (which equals 153) are formed, that is, 0–153 Hz, 154–306 Hz, and 307–459 Hz. Accordingly, these three pitch range types are coded as small, medium, and large, respectively. This mechanical classification procedure is not based on the pitch range perception. However, with this classification, the relation between pitch range value and pitch range type of the vocal-verbal micro-feedback and understanding can be investigated in this study.

#### 4.1.10 Reliability control of the data

The reliability of the research data (for different empirical studies) is controlled in different ways as follows.

##### On micro-feedback and understanding annotation

In order to identify micro-feedback and annotate whether it conveys sufficient understanding, misunderstanding, or non-understanding, inter- and intra-coder reliability checking has been carried out between five Chinese and seven Swedish transcribers and annotators in total. Five of them were the main transcribers and annotators who did the majority of the work, and the rest also did the work to varying degrees. First, one main transcriber (who is also the author) was trained in the transcription standard and the coding schemes, and she transcribed and annotated the first five minutes of each conversation. All the other transcribers and annotators were trained to transcribe and annotate the audio and video materials using the same transcription standard and coding schemes, primarily the GTS and the MUMIN. In addition, one Chinese and two Swedish transcribers and annotators worked together with a sample of one hundred occurrences in order to establish a common procedure that could be used by all the transcribers and annotators. As a departure from what the main transcriber had done with the first five minutes of each conversation, the other transcribers, who were also the corresponding annotators, sequentially checked and corrected the first five minutes' transcriptions and annotations. After

---

in 1987. More information about the INSTINT system can be found at <http://www.lpl.univ-aix.fr/~hirst/intsint.html>.

that, they continued to transcribe and annotate in full each conversation. The person involved in the project work was both the transcriber and annotator of the conversation. In total, each interaction was transcribed and annotated by at least two persons (including the author for the first five minutes and the other person for the remaining time) and then checked by another two persons. During the process, inter-coder (between different transcribers and annotators) and intra-coder (within the transcriber and annotator) reliability checking had been carried out continuously. In the final joint reliability checking process, two of the transcribers and annotators together checked the transcribed and annotated data with a randomly selected sample of one hundred instances of micro-feedback from different dialogues. The purpose was to examine how much in general the different transcribers and annotators involved in the project could agree on as regards the completed transcriptions and annotations. The result shows that the agreement rate is 93 out of 100. The data were thus regarded as sufficiently reliable.

Self-confrontation interviews were also carried out when the transcribers and annotators had difficulties in identifying or interpreting the communicative functions of the micro-feedback. For instance, when the annotators were not certain about whether the participants understood or not, whether the participants were communicating agreement or just signalling perception and understanding, or whether the participants were surprised or excited, self-confrontation interviews were conducted. Not all the participants were asked to participate in this self-confrontation interview, only the ones who were involved in the conversations that were difficult for us to transcribe and annotate. Besides, not the entire interaction was presented to the participants, only the dialogue segment that was difficult to transcribe and annotate. In the interviews, the participants were provided with a film of a longer context than the dialogue segment. No more content than that was shown to the participant interviewed, unless the participant made a request to see more in order to achieve self-confrontation. In this way, some of the interlocutor's intentionality and understanding during the interaction was justified in the annotation. One unanticipated consequence of the work that had been done was that people's self-recognition was enhanced and the relevant coding scheme was enlarged to some extent.

#### Cohen's kappa inter-coder rating of pitch contour annotation

As presented in Section 4.1.9, based on Tronnier and Allwood (2004) and Cerrato (2005), a simpler set of criteria is used to code the pitch contour in this study, which includes rising, flat, and falling. The prosodic segmentation and annotation were carried out by one annotator, and inter-coder reliability checking was performed by another independent annotator separately. First, these two annotators (one Chinese

and one Swedish) were trained by using the same coding scheme for pitch contour annotation (i.e., three categories: rising, flat, and falling). With the same criteria, these two annotators practised together on roughly 30 instances. Then, the two annotators coded the data independently. One annotator coded the entire data (i.e., 661 vocal-verbal micro-feedback sound clips), whereas the other coded a random selection of 20% of the entire data (i.e., 132 instances). Each random number of the selected clip was generated via the website [www.random.org](http://www.random.org). Following this, the inter-coder agreement was rated by Cohen's kappa<sup>39</sup> (Cohen, 1960). The result shows that the percent of overall agreement  $P_o$  (i.e., percent agreement observed) is 0.81, and the free-marginal kappa (see footnote for Cohen's kappa) is 0.72.

Although many linguistic and communication researchers do not fully agree on how to measure inter-coder agreement in annotated data and how to interpret the result (see Artstein & Poesio, 2008), among others Fleiss (1971) suggested that Cohen's kappa figure over 0.60 was good and over 0.75 was excellent. Landis and Koch (1977) characterised Cohen's kappa values < 0 as indicating no agreement and 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. Also, according to Brennan and Prediger (1981), kappa values can range from –1.0 to 1.0, with –1.0 indicating perfect disagreement below chance, 0.0 indicating agreement equal to chance, and 1.0 indicating perfect agreement above chance; a rule of thumb is that a kappa of 0.70 or above indicates adequate inter-rater (or inter-coder) agreement. Thus, the Cohen's kappa result of 0.72 obtained in this study is sufficiently reliable to conduct further analysis.

#### 4.1.11 SPSS statistical tests on the prosodic data

In order to test the association between prosody (which includes pitch, pitch contour, pitch range values and types, and duration values and types) of the micro-feedback and understanding, the descriptive statistical research method box plot and the inferential statistical methods Kruskal-Wallis statistical test, Fisher's Exact Test, and ANOVA statistics are used in the present study.

These methods are chosen, in accordance with Howell (2010). Box plotting is used to present the distributions of the pitch values (including minimum pitch, maximum pitch, and mean pitch), the pitch range values, and the duration values.

---

<sup>39</sup> Two variations of kappa are available: Siegel and Castellan's (1988) fixed-marginal multi rater kappa and Randolph's free-marginal multi rater kappa (see Randolph, 2005; Warrens, 2010). Brennan and Prediger (1981) suggest using free-marginal kappa when raters are not forced to assign a certain number of cases to each category and using fixed-marginal kappa when they are. In this study, free-marginal kappa is applied with an online tool on the webpage <http://justusrandolph.net/kappa/>.

Because the distributions of the studied data for pitch and duration are non-normal (i.e., the distribution of pitch is bimodal and that of pitch range values and duration values are positively skewed), the Kruskal-Wallis one-way analysis of variance is applied in order to examine the associations between the pitch values, the pitch range values, or the duration values and the three different types of understandings (i.e., sufficient understanding, misunderstanding, and non-understanding).

The duration values (in milliseconds) are, in addition, transformed into a normal distribution by taking the logarithm of the values, and then tested by ANOVA to investigate its relation to understanding.

The Fisher's Exact Test is used to test the associations between the pitch contour, the duration type, or the pitch range type and understanding, and also the associations among pitch contour, duration type, and pitch range type. Fisher's Exact Test was chosen over the Chi-square statistical test, because both of them function equally well in testing the significance of frequency differences and the Chi-square statistical test deals with in particular large frequencies whereas the Fisher's Exact Test can handle both large and small frequencies.

Furthermore, Fisher's Exact Test  $3 \times 3$  (two-tailed) is set up to examine whether there is any or no association in the data studied. The reason for the two-tailed part of the test is that it is not known beforehand which frequencies would be lower and which ones would be higher. Both directions are considered equally likely (or unlikely). This type of statistical test is commonly used for explorative studies and logically inductive research like the present study. Following this, if there is some kind of association, the pairwise comparisons with the Fisher's Exact Test  $1 \times 2$  (one-tailed) is used to examine the details of this association by testing for statistical differences between the observed frequencies. For instance, it is used to investigate how the studied contents are associated exactly, when it can be assumed that one direction (e.g., non-understanding is expressed more in a rising pitch contour than falling, or non-understanding is expressed less with falling than rising pitch contour) is considered more extreme than the other.

In statistical significance testing, the one-tailed test and the two-tailed test are alternative ways of computing the statistical significance. According to Howell (2010), one-tailed and two-tailed tests in Fisher's Exact Test are not clear-cut. Instead of going into how to determine what is a more extreme outcome, Howell (2010, p. 148) suggested avoiding that complication and simply making a decision in advance whether to use a one- or a two-tailed test and then reporting the values computed by the standard statistical software.

## 4.2 Method of Study 2

In order to study understanding further, for instance, to reconceptualise the notion of understanding and investigate how understanding problems are detected, handled, and resolved in sense-making and information sharing, the empirical data was expanded in Study 2 (see more reasons stated in Chapters 1 and 8).

Many researchers have claimed that context and all the relevant constraints of the communicative situation and communicative task influence language use, the discourse content, and the linguistic behaviours in communication (see e.g., Finkbeiner et al., 2012). So, as a contrast to the earlier data in Study 1, which focuses on the simple task of getting acquainted in FTF first encounters, the data in Study 2 focuses on complex task-solving collaboration in both FTF and VMC. Data for Study 2 were collected from ten FTF and ten VMC acquaintances' conversations between ten Swedish and ten Chinese speakers who were instructed to jointly solve some learning assignments with reference to reading material provided.

This data has been studied in three empirical analyses which result in a contribution to theoretical understanding of understanding in human interaction. The first analysis focuses on the conceptualisation of understanding in interaction. The second analysis investigates how understanding problems are handled by interlocutors. The third analysis compares understanding and understanding problems between FTF and VMC. Many different levels and stages of understanding occur in information sharing, but only the overt (revealed and manifested) understandings are studied in this thesis.

The interactional approach applied and the basic research design issues that are similar between Study 1 and 2 are not repeated here, for example, intercultural communication, dyadic conversation, communication language, audio and video data transcription standard, and micro-feedback annotation and coding scheme. Only the new methodological issues and designs for Study 2 are presented here.

### 4.2.1 Video-mediated communication (VMC)

In order to provide more variety of the data and higher reliability of the research on understanding in interaction, the research project is expanded by including both FTF and VMC interactions in Study 2. This not only complements some research limitations in Study 1, but also provides useful insights for future information and communication technology (ICT), in particular video related interaction innovations and applications. As presented, the main reason for enriching the data with VMC interactions in Study 2 is that I anticipate a larger number of understanding problems than in FTF and that more reliable conclusions regarding understanding and its

problems in interaction will be generated. Understanding and understanding problems will be studied in both FTF and VMC; whereas, FTF versus VMC is not the main topic of this thesis<sup>40</sup>. Although VMC has different modes such as text chat or messaging, only the audio and video chat of VMC, namely the auditory vocal-verbal and visual gestural communication, is studied in this thesis. Spontaneous dyadic interaction is chosen for the same reasons as stated in Chapter 4.

#### 4.2.2 Between university students who were acquainted

Understanding between people of different interpersonal familiarity has been regarded as important as understanding between different cultures. Navarretta and Paggio (2013) and Campbell (2007) have stressed that among other factors such as the physical setting, the number of participants, and the topic discussed, the degree of familiarity influences a lot the use of micro-feedback and the mutual understanding in interaction. Baltes, Dickson, Sherman, Bauer, and LaGanke (2002) and Stone and Posey (2008) have also claimed that familiarity affects norms of behaviour and effectiveness of coordination, cohesion, interaction, and understanding in communication.

According to Svennevig (1999, pp. 1–6), interpersonal relationships consist of three constitutive components: (1) solidarity, involving a set of mutual rights and obligations; (2) familiarity, involving mutual knowledge of personal information; (3) affect, involving mutual liking (or disliking). The first encounter (Study 1) is characterised primarily by the first two components, whereas deeper relationships such as friendship and partnership include also the third. The collaborative task-solving data in Study 2, for example, involves this third component in that the other participant for the joint task collaboration was deliberately chosen by the initiating participant who volunteered to participate in the project. The data in Study 2 provides deeper interpersonal relationships than that of Study 1.

Kiesler and Sproull (1992) argued that unacquainted people had an advantage in that individuals had reduced social context cues and overcame social inhibitions thus leading to more equal participations. However, unacquainted people have different individual communication presuppositions and expectations as well as various limitations in common knowledge and resources in sense-making and information sharing (Linell, 2009). It is likely more difficult for the unacquainted people who have mutually distinct and unknown personal and professional experiences to achieve mutual understanding than it is for acquaintances. However, Ermakova and

---

<sup>40</sup> How people communicate and co-construct sense-making and information sharing through micro-feedback in FTF and VMC is studied only in Chapter 10 in Study 2 in this thesis.

Zemskaja (1993) and Sarangi (1994) (cf. Mustajoki, 2012) have pointed out that acquainted people may have similar communication failures and understanding problems as do unacquainted people, depending on the discourse type and communication context (see more discussion in the next section). A variety of communication situations are needed for the study of understanding and communication.

Risks of miscommunication are very genre-specific. The genre (discourse type) determines not so much the frequency, but chiefly the type of communication failures. There are interesting observations on this. On the basis of a large authentic material, Ermakova and Zemskaja (1993) have reached a striking conclusion: communication failures are as usual between good friends and relatives as they are in intercultural encounters (cf. Sarangi, 1994). However, the risks and causes of miscommunication in these genres are very different. We need further studies of a variety of communication situations to reach a more complete understanding of the role of the genre factor in cases of miscommunication (Mustajoki, 2012, p. 237).

With the aim of supplementing a variety of the data and counterbalancing the restriction facing the unacquainted participants in Study 1, acquainted participants were recruited and studied in Study 2 in this thesis. The participants were ten Swedish (three female and seven male) and ten Chinese (seven female and three male) university students, who were already acquainted or friends with one another, were 24 years old on average, and were studying in Sweden. They had higher interpersonal familiarity than the participants (who were strangers) in Study 1. The communication language in Study 2 remained the same as that in Study 1, that is, the participants used English lingua franca (see discussions earlier in Section 4.1.5) in the conversations. The fact that all the participants are non-native English speakers may influence how they communicate with and understand each other. Although second language usage and its influence on communication is not the focus of the present study, how much of an understanding problem is associated with this will be discussed in the empirical analyses chapters and the conclusion and discussion chapters later.

### 4.2.3 A complex communication task

Communication and understanding behaviours may vary between simple and complex communication contexts.

Hancock and Dunham (2001) have observed that people have more difficulties completing a joint problem-solving task and they often have to use more meta-communicative signals, such as micro-feedback, to achieve their interaction goals. Also, Lindwall et al. (2005) and Sins et al. (2011) have found that because a successful

learning cooperation requires intensive high-quality discourse exchange, understanding is crucial to solving the learning assignments and thus achieving successful learning cooperation.

Study 2 in this thesis is based on a complex communication context of *task-solving collaboration* to solve some learning assignments. The learning assignments used in the project are based on reading material<sup>41</sup> about cultural frameworks and their importance for leadership and communication (see Appendix C), which was provided by the researcher at the beginning. Each pair of participants was given two tasks to discover similarities and differences between China and Sweden, preferably within eight minutes.

This particular reading material was chosen because of a few reasons. First, a topic like such of cultural awareness, especially in the globalising world, is interesting for the participants to engage in reading, thinking, and discussing. Second, this material is relevant and suitable for the participants, because they are acquaintances or classmates coming from different cultural backgrounds and it is natural that they are curious and want to develop the knowledge of each other. Third, this topic and the designed questions are open ended, which allow and encourage communication, discussion, and negotiation.

According to Anderson (2006, p. 258), “with experience in problem-solving tasks, speakers become more efficient communicators in a number of ways. For example, over time, speakers use shorter referring expressions, need to say less to complete the task, and coordinate more effectively... at establishing common ground”. The two tasks were counterbalanced in order between the FTF and VMC communication situations. Also, open discussion questions comprise the communication tasks in order to help the participants learn the context of the material, shape their conversation, and uncover understanding and understanding problems during conversation. However, the participants’ understanding of the provided reading material itself and the involved terminologies are not the focus of the study.

#### 4.2.4 Audio- and video-recorded dialogue data

Both primary and secondary data were collected with audio- and video-recording apparatuses in Study 2.

---

<sup>41</sup> The reading material is an extract from Chapter 3 in the book “*Leadership in a Diverse and Multicultural Environment: Developing Awareness, Knowledge, and Skills*” (Connerley & Pedersen, 2005). The book is mainly used for undergraduate- or graduate-level international management, leadership, or diversity-related courses taught in the business curriculum, and it is also used in leadership courses taught in the education and communication departments at the University of California.

### Primary data from audio and video FTF and VMC interactions

In the FTF situation, participants were audio- and video-recorded in a lab with an appearance of a normal study room, equipped with a study desk, a chair, a pen and the reading material (see Figure 4.8). The participants communicated while sitting as they normally do in real learning groupwork. Also, because hand and lower body movements are practically not involved in micro-feedback-related understanding behaviours (Lu, 2014; Navarretta & Lis, 2012), this set-up does not affect the research purpose of the present study. One microphone for each participant was used for the audio recording. Three cameras (positioned left, centre, and right) were used for the video recording.



Figure 4.8. Task-solving collaboration in FTF (two pairs of participants).



Figure 4.9. Task-solving collaboration in VMC (two pairs of participants). (Note: the bottom image shows what the participants saw on their screens.)

In the VMC situation, two rooms that are identical in size and appearance were used (i.e., one participant in each). Participants communicated with one another from their own room. The participant sat in front of a study desk with a 24" computer monitor and a mouse, which were identical for both participants. For each participant, the same microphone as used in FTF was used in VMC. Interaction in VMC

was achieved by using the software Skype. The participant was video recorded by both the automatic webcam which was filming straight at the participant (with a focus on the top of the upper body) and an extra camera situated at the side filming at an angle of approximately 45° (filming the whole upper body) (see Figure 4.9).

The webcam image of the other participant was enlarged to fill the participant's monitor. Factors affecting the quality of speech and interfering with the process of communication were minimised. A manual intercom system was used to make it possible for both the researcher and the participants to inform each other when to start and end the activity and recording.

The video-recordings last between 4:28 and 15:57 minutes (with a mean of 9:28), consisting of 472 to 1,744 vocal words (with a mean of 1,033). In all, the recordings last 3:09:28 minutes and consist of 20,669 vocal words (see Table 4.2). Each entire interaction is analysed in the present study.

*Table 4.2.* The size of the analysed data in Study 2.

<b>Dialogue</b>	<b>Communication situation</b>	<b>No. of words</b>	<b>Time length (min.: sec.)</b>
1	VMC	752	06:27
2	FTF	1110	09:35
3	FTF	1377	12:35
4	VMC	965	08:18
5	VMC	1327	08:41
6	FTF	741	04:28
7	FTF	1165	08:52
8	VMC	1177	09:30
9	VMC	842	11:45
10	FTF	472	08:33
11	FTF	1744	12:59
12	VMC	1479	09:50
13	VMC	1459	11:20
14	FTF	874	07:08
15	FTF	804	08:42
16	VMC	525	05:04
17	VMC	1059	10:39
18	FTF	539	05:20
19	FTF	1028	13:45
20	VMC	1230	15:57
Total		20669	3:09:28

## Secondary data from interviews and questionnaires

As proposed by Mustajoki (2012, p. 236), “for a proper analysis of the course of dialogues”, previous dialogue contexts between the interlocutors should be taken into account and “perhaps indeed their whole lives prior to the moment of communication” and their subsequent feedback as well. Full implementation of these requirements was not realistic in my study. However, a number of secondary data were collected additionally. A general personal information questionnaire on the participant’s national and educational background (see Appendix D) was conducted beforehand. An individual follow-up interview on the content of what the participant had learned from the reading material and from the task-solving collaboration was held and an individual questionnaire on the communication experience was given straight afterwards sequentially (see Appendix E and F). The interviews were audio recorded.

In the analysis process, the secondary data from the follow-up interviews and questionnaires were used to help interpret and code the participants’ understanding and communication behaviours, in conjunction with the primary data from audio- and video-recorded dialogues. Because the participants were instructed to solve some learning assignments, they mainly worked together to discuss and arrive at solutions. Whether and how they understood each other was revealed in the primary and secondary data. Consequently, self-confrontation interviews, which were used in Study 1 were not employed in Study 2.

### 4.2.5 Independent and dependent variables in the comparative study

In the comparative study in Chapter 10, in particular, independent and dependent variables are measured and the results are discussed in relation to the earlier theories (presented mainly in Chapters 2 and 3).

The independent variable of the communication situation is compared, that is, FTF and VMC (mainly in Chapter 10). Dependent variables of understanding, that is, sufficient understanding, misunderstanding, non-understanding, and other analytical results concerning understanding and understanding problems (e.g., partial understanding and meaning repair in terms of its initiation and performance (presented in Section 2.9)) are coded and assessed.

#### 4.2.6 Data collection procedure

At first, one pilot study was carried out with two pioneer participants. Then, some details of timing and apparatus settings were refined and settled. The data were collected in the following procedure.

Two participants in the same pair were, first, given a personal information questionnaire, one each in separate rooms, where the participation was initiated. Once they had completed the questionnaire, the participants were provided with the academic material to read in separate rooms. After having read the article, the two participants were led to the corresponding communication room(s) (one common room for FTF, two separate rooms for VMC). Both the communication tasks and the communication situations are counterbalanced in this study. That is, in the first communication situation (FTF or VMC), the participants were given open discussion questions 1 (i.e., three similarities/differences between China and Sweden and point out the most obvious one) (see Appendix G). Then, in the second communication situation (VMC or FTF), the participants were given the discussion question 2 (i.e., three differences/similarities between China and Sweden and point out the most obvious one). The chronological orders of communication media and communication tasks were counterbalanced. The length of each interaction was recommended to be approximately eight minutes, although it primarily depended on how long it took the participants to complete the task. An overview of the size of the data in Study 2 is presented in Table 4.3.

After having communicated in two situations (FTF and VMC), the participants were led to the two separate rooms where they had begun in the project earlier. There, they were given an individual follow-up interview on the content of the reading material and the result of their discussion. The purpose was to get the participant to reflect on what (s)he had learned from this collaborative learning activity and also what and how (s)he had understood from the other participant. After this, there was an individual interview on the experience of communication media and the difficulty of understanding. At the end, the participant answered a questionnaire on his/her personal attitude towards VMC and FTF, self-assessment of the interactions, and individual communication experience as a whole.

Table 4.3. Overview of the data in size (Time = hour:minute:second; Word/ Utter. = number of words/ utterances; Dialogues are numbered according to the time when they were recorded; Dialogues are named as follows: first the dialogue number, then the communication task (d = differences and s = similarities) and the recording date and last the communication media (v = vmc and f = ftf)).

VMC data				FTF data			
Dialogue number	Time	Word	Utter.	Dialogue number	Time	Word	Utter.
D1d(150121).v	06:27	752	121	D2s(150121).f	09:35	1110	179
D4s(150123).v	08:18	965	133	D3d(150123).f	12:35	1377	226
D5s(150204).v	08:41	1327	159	D6d(150204).f	04:28	741	103
D8d(150209).v	09:30	1177	134	D7s(150209).f	08:52	1165	186
D9d(150218).v	11:45	842	110	D10s(150218).f	08:33	472	62
D12s(150220m).v	09:50	1479	197	D11d(150220m).f	12:59	1744	257
D13s(150220e).v	11:20	1459	192	D14d(150220e).f	07:08	874	148
D16d(150309).v	05:04	525	100	D15s(150309).f	08:42	804	161
D17d(150310).v	10:39	1059	143	D18s(150310).f	05:20	539	91
D20s(150314).v	15:57	1230	157	D19d(150314).f	13:45	1028	160
Subtotal of VMC	1:37:31	10815	1446	Subtotal of FTF	1:31:57	9854	1573
VMC per minute	97.52	110.90	14.83	FTF per minute	91.95	107.17	17.11
Entire total	3:09:28	20669	3019	Entire per minute	189.47	109.09	15.93

#### 4.2.7 Transcription and annotation

In this thesis, Study 2 was an expansion and continuation of Study 1. Consequently, the primary technical systems used to carry out the analyses remained more or less the same. The recorded data were manually transcribed and annotated according to variants of The Göteborg Transcription Standard (henceforth GTS) version 6.4 (Nivre et al., 2004) and the MUMIN coding scheme (Allwood et al., 2007). Since Study 2 is relatively more focused on the context of interaction and understanding than micro-feedback and its particular linguistic features, it is transcribed and annotated slightly differently from Study 1. Study 1 was transcribed and annotated with a focus on micro-feedback (e.g., detailed vocal-verbal, gestural, and emotional annotations of every single micro-feedback expression), and Study 2 was transcribed and annotated with a focus on understanding (e.g., necessary and sufficient vocal-verbal, gestural, and emotional annotations of micro-feedback that are of relevance to understanding).

Schemes for annotating understanding from Study 1, including sufficient understanding, misunderstanding, and non-understanding, together with the analytical finding of partial understanding from Study 2 are employed in the analysis. The term partial understanding has been used by a few researchers (e.g., Allwood, 1986), to refer to some type of understanding problem or difficulty. Briefly, partial understanding occurs when the interlocutor can only partially but not sufficiently make sense or share the meaning of the information presented and the interaction does not continue as intended and anticipated. Partial understanding is a developing stage towards achieving sufficient understanding.

#### 4.2.8 Reliability control of Study 2: Cohen's $\kappa$ inter-coder rating on understanding annotation

As presented above, sufficient understanding, misunderstanding, and non-understanding from Study 1 together with the analytical result of partial understanding from Study 2 are employed to annotate understanding in the study. The annotation was carried out by one annotator, and inter-coder reliability checking was performed by another annotator independently. First, these two annotators (one Chinese and one Swedish) were trained by using the same coding scheme for understanding annotation (i.e., four categories: sufficient understanding, misunderstanding, non-understanding, and partial understanding). With the same criteria, these two annotators practised together on 10 instances. Then, the two annotators coded the data separately. One annotator coded the entire data (i.e., 2204 occurrences of different understandings). The other coded a random selection of 20 instances of each type of understanding, that is, 80 instances of understandings in the entire data. Twenty cases of each type of understanding were selected, because there were in total 21 instances of misunderstanding found in the data. Each random number of the selected clip was generated via the website [www.random.org](http://www.random.org). Following this, inter-coder agreement was rated by Cohen's kappa<sup>42</sup> (Cohen, 1960). The result shows that the percent of overall agreement  $P_o$ . (i.e., percent agreement observed) was 0.69, and the free-marginal  $\kappa = 0.58$ . According to the suggestions made by Fleiss (1971), Landis and Koch (1977), and Brennan and Prediger (1981) on how to evaluate the result (see details of scales presented in Section 4.1.10), the inter-coder agreement on the annotation of understanding is moderate.

---

<sup>42</sup> Two variations of kappa are available: Siegel and Castellan's (1988) fixed-marginal multi-rater kappa and Randolph's free-marginal multi-rater kappa (see Randolph, 2005; Warrens, 2010). Brennan and Prediger (1981) suggest using free-marginal kappa when raters are not forced to assign a certain number of cases to each category and using fixed-marginal kappa when they are. In this study, free-marginal kappa was applied.

#### 4.2.9 Coding for understanding and understanding detecting and resolving

As mentioned earlier, Study 2 is a development from Study 1 and contributes to a more nuanced classification of understanding, namely partial understanding. Thus, understanding in Study 2 is categorised into sufficient understanding, misunderstanding, non-understanding, and partial understanding. This classification is part of the analytical results of the conceptual analysis of understanding in Chapter 8. More details of the definition and operationalisation criteria will be presented there.

Briefly, in addition to the earlier coding scheme of understanding presented in Section 4.1.9 (not to be repeated here), partial understanding is identified as a form of insufficient understanding or understanding problem. It takes place when the interlocutor can only partially but not sufficiently make sense or share the meaning of the information presented, and the interaction does not continue as intended and anticipated. Partial understanding is a developing stage towards achieving sufficient understanding, and it usually succeeds eventually. Therefore, accordingly, the data focused on were coded in terms of sufficient understanding, misunderstanding, non-understanding, and partial understanding.

Regarding how to analyse the detection and the resolving of the identified understanding problems, a few trial and learning data sessions were conducted in collaboration with other researchers. Coding schemes for analysing the material concerning detected versus undetected and resolved versus unresolved understanding problems were decided on, with regard to whether the understanding problem is detected and resolved by the participants during the interaction. The empirical analyses were conducted from an analytical perspective.

As discussed earlier in the theory chapter (Chapter 2), identifying whether the information is understood or not is a methodologically difficult issue (see Sections 2.3 to 2.8). However, because this abstract phenomenon of understanding is manifested through communicative behaviours, interpretation of understanding can be carried out through introspection or observation by using evaluation criteria on contextualisation, relevance, meaning and implicature, and activity type (see earlier theoretical discussions). The most common approach is to use an analytical perspective. The analyst's interpretations can provide insights that the persons involved (i.e., participants) probably do not see. Tests of inter-rater reliability can make this less subjective.

In addition, follow-up interviews were conducted with individual participants immediately after their conversations (see details in Appendix E), in order to verify a few issues for the analysis: what the participants (believed themselves) had understood from the tasks, how they had solved them, what they had understood from

their partners, what they had agreed or disagreed about with each other, what they felt to be problematic during the task-solving progress (in terms of both content (i.e., tasks and topics), context (i.e., FTF and VMC), and collaboration (i.e., interpersonal and group work)), and what they had learned from the reading materials. How much the participants could recall and reflect on their participation in the project remains a question, however, these follow-up self-reflection interviews provided useful information to help the coders to justify the analytical interpretations.

#### 4.2.10 SPSS statistical tests on differences between FTF and VMC

In order to see which quantitative differences in understanding and understanding problems between FTF and VMC are statistically significant, a number of statistical tests were carried out using IBM SPSS Statistics 25. All tests were two-tailed and alpha was set to .05. Such quantitative comparisons have been applied in many studies of FTF versus VMC or CMC, such as Straus and McGrath (1994), Anderson (2006), and Stone and Posey (2008).

The conversations from VMC and FTF were compared concerning length, number of words, and number of utterances. Since the participants and dyads participated in both VMC and FTF, and since there were two conditions to be compared, a repeated measures *t* test was used in order to see if the means of the two conditions differed. The sample size is somewhat small ( $N = 10$ ) because each dyad generates extensive interaction data that requires considerable resources to transcribe and analyse. This transcribed data was the main interest rather than the comparisons at the level of dyads. Although the sample sizes are small, the *t* test can detect if there are any large differences, which is not unreasonable to expect (e.g., task completion times could differ greatly between VMC or FTF).

Whether certain types of understanding were connected to VMC or FTF was tested by looking at the frequencies of understanding. Each occurrence of understanding was considered a data point. A Chi-square test of independence was carried out to test whether the categorical variable communication situation (VMC and FTF) was associated with the categorical variable understanding (sufficient understanding, misunderstanding, non-understanding, and partial understanding). Frequencies were sufficiently high for this test to be suitable (Howell, 2010). Here, two tests were performed, one featuring the raw frequencies and one featuring frequencies per time unit. The latter was reformulated into occurrence per 100 minutes in order to yield integers for the Chi-square test (so that, e.g., 0.34 occurrences per minute became 34 occurrence per 100 minutes).

### 4.3 Ethical consideration

Participants were recruited by means of an advertisement. They voluntarily contacted the researcher and participated in the project. All of them were over 18 year of age and regular university students. The average age of the participants in Study 1 was 26, and in Study 2 the average was 24. They gave individual written consent to participate. The consent form for participation in the project is presented in Appendix H. The research in this thesis is not considered to uncover information that requires review by the Regional Ethical Review Board of Gothenburg. The material does not include sensitive personal data. However, the thesis work was conducted with ethical considerations of not violating participant's consent as regards, for example, their personal information, the use of this research data, access to it, and its publication. The participants gave their consent to being audio and video recorded, and that their data could be used only for academic purposes. The participants had control over how much or little they took part in the research project. They could choose not to respond to the questionnaires or the interviews or not let the researcher use their images for publication. However, in the research process, the participants showed clearly active and positive participation in such academic research. The participants were anonymised in both the analyses of the data and the presentations of the results. The audio and video research data were stored on a secure hard drive, which has restricted accessibility. The participants were assured of confidentiality. This research project received approval from the Department of Applied Information Technology at the University of Gothenburg before it was carried out.



## Results of modality analysis

This chapter presents an empirical analysis, in which micro-feedback is investigated in detail with respect to its modality in relation to sufficient understanding, misunderstanding, and non-understanding. The vocal-verbal and gestural micro-feedback studied in this thesis is distinguished in terms of sensory modality, that is, auditory modality and visual modality.

Two research questions will be investigated. First, how are the auditory and visual modalities involved in micro-feedback expressions that are related to sufficient understanding, misunderstanding, and non-understanding? Second, what are the typical unimodal and multimodal micro-feedback expressions that signal sufficient understanding, misunderstanding, and non-understanding?

### 5.1 Overview of the data studied

As discussed earlier, the research data in Study 1 consist of eight Swedish–Chinese intercultural communication dialogues between four Swedish and four Chinese speakers, who were on average 26 years old and studying at universities in Sweden. The dialogues last 65 minutes and 8 seconds in total and consist of 10,127 vocal words. The data were transcribed, annotated, and checked according to GTS and a variant of MUMIN. The different understandings were annotated according to the classification of sufficient understanding, misunderstanding, and non-understanding.

Table 5.1. Unimodal and multimodal micro-feedback in relation to sufficient understanding, misunderstanding, and non-understanding (*Note.* Numbers are frequencies<sup>43</sup> (abbreviated as F.) per 1,000 words and per minute. Suff. = sufficient).

Micro-feedback	Suff. understanding			Misunderstanding			Non-understanding			Total		
	F.	/1,000 words	/min.	F.	/1,000 words	/min.	F.	/1,000 words	/min.	F.	/1,000 words	/min.
Unimodal vocal-verbal	336	33.18	5.16	3	0.3	0.05	2	0.18	0.03	341	33.68	5.24
Unimodal gestural	341	33.68	5.24	0	0	0	2	0.18	0.03	343	33.87	5.26
Unimodal total	677	66.86	10.4	3	0.3	0.05	4	0.36	0.06	684	67.55	10.5
Multimodal total	579	57.17	8.89	6	0.59	0.09	19	1.88	0.29	604	59.65	9.28
Total	1,256	124	19.28	9	0.89	0.14	23	2.27	0.35	1,288	127.18	19.78

As presented in Table 5.1, the participants used more unimodal micro-feedback expressions than multimodal ones (as frequencies of 10.5 and 9.28 per minute) to express the different understandings. Sufficient understanding is signalled more through the unimodal micro-feedback expressions than the multimodal ones (with a frequency of 677 compared to 579). Misunderstanding involves more multimodal micro-feedback than unimodal (6 to 3). Most of the non-understandings are expressed by multimodal micro-feedback than unimodal (19 versus 4).

The participants conveyed overwhelmingly sufficient understanding rather than misunderstanding and non-understanding through micro-feedback, with a high frequency of 1,256 out of a total of 1,288 (see Table 5.1). There are only small numbers of misunderstanding and non-understanding (9 and 23, respectively). Details of how sufficient understanding, misunderstanding, and non-understanding are conveyed by means of which specific unimodal and multimodal micro-feedback will be investigated in the following sections.

## 5.2 Unimodal micro-feedback

As shown in Table 5.1, the participants used the unimodal vocal-verbal micro-feedback and the unimodal gestural micro-feedback roughly equally often to express the different understandings, with frequencies of 5.24 and 5.26 per minute. The unimodal vocal-verbal and unimodal gestural micro-feedback expressions in relation to the different understandings will be studied in the following sections.

<sup>43</sup> In this thesis, frequencies that are not whole numbers are rounded to two decimal places.

## 5.2.1 Unimodal vocal-verbal micro-feedback

In this section, how sufficient understanding, misunderstanding, and non-understanding are communicated through the unimodal vocal-verbal micro-feedback will be sequentially presented.

Table 5.2. Overview of the unimodal vocal-verbal micro-feedback related to sufficient understanding, misunderstanding, and non-understanding.

Micro-feedback	Participant	Sufficient understanding	Mis-understanding	Non-understanding	Total
Unimodal vocal-verbal	Swedish	138	0	0	138
	Chinese	198	3	2	203
	Total	336	3	2	341

The overview of unimodal vocal-verbal micro-feedback presented in Table 5.2 shows that, in the data, only the Chinese participants used unimodal vocal-verbal micro-feedback to communicate that they did not understand or that they misunderstood the information. In the data, the Swedish participants used unimodal vocal-verbal micro-feedback only to express sufficient understanding. The Swedes did not use any unimodal vocal-verbal micro-feedback related to misunderstanding or non-understanding. Unimodal vocal-verbal micro-feedback in relation to sufficient understanding will be investigated first, and later in relation to misunderstanding and non-understanding, respectively.

### For sufficient understanding

In the data, the most common unimodal vocal-verbal micro-feedback used by the participants to signal sufficient understanding are words *yeah*, *okay*, *m*, *ah*, *yes*, and *no* (see Table 5.3).

Table 5.3. The most common unimodal vocal-verbal micro-feedback expressions (VFB) that are used to signal sufficient understanding (F. = frequency).

<b>Unimodal VFB</b>	<b>F.</b>	<b>/1,000words</b>	<b>/min.</b>
<i>yeah</i>	95	9.39	1.46
<i>okay</i>	40	3.95	0.61
<i>m</i>	31	3.06	0.48
<i>ah</i>	11	1.08	0.17
<i>yes</i>	9	0.89	0.14
<i>no</i>	8	0.79	0.13
<i>uhu</i>	7	0.69	0.11
<i>yeah yeah yeah</i>	6	0.59	0.09
<i>oh</i>	5	0.5	0.08
<i>m:</i>	4	0.4	0.06
<i>aha</i>	4	0.39	0.06
<i>yeah yeah</i>	4	0.39	0.06
<i>ah yeah</i>	4	0.39	0.06
<i>yeah okay</i>	4	0.39	0.06
<i>ah okay</i>	3	0.3	0.05
<i>mhm</i>	3	0.3	0.05
<i>okay okay</i>	3	0.3	0.05
<i>cool</i>	2	0.2	0.03
<i>o:kay</i>	2	0.2	0.03
<i>eh which part</i>	2	0.2	0.03
<i>it's a big city</i>	2	0.2	0.03
<i>gym</i>	2	0.2	0.03
<i>sandra ah</i> (person's name)	2	0.2	0.03
Others (F. = 1)	83	8.18	1.27
Total	336	33.18	5.17

The unimodal vocal-verbal micro-feedback words *yeah*, *okay*, and *m* are mostly used to express “I hear you, I perceive and understand what you have said, and I would like to continue the conversation” as annotated with CPU (see Excerpts 1 to 3).

**Excerpt<sup>44</sup> 1: Unimodal vocal-verbal micro-feedback *yeah* (from Dial. 1)** (Note. The key information is bolded.)

\$Cm1: that's gonna be cool yeah because i prefer the drawing when i // study in the uh / uh senior school  
... ..<sup>45</sup>  
\$Sm1: < yeah >  
@ < VFB; CPUE/A agreement >, < GFB; head: nod; CPUE/A agreement >  
\$Sm1: < **yeah** >  
@ < **VFB; CPU** >

**Excerpt 2: Unimodal micro-feedback *okay* (from Dial. 2)**

\$Cm1: karin / oh yeah karin okay / are are you a swe+ swedish ... ..  
\$Sf1: < yeah i'm swedish >  
@ < GFB head: nods; CPUE/A confirmation >, < GFB general face: smile; CPUE/A friendliness/awkwardness >, < VFB; CPUE/A confirmation >  
\$Cm1: < **okay** > so de+ eh city one is your home town  
@ <1 **VFB; CPU** >1

**Excerpt 3: Unimodal micro-feedback *m* (from Dial. 2)**

\$Sf1: < yeah > well we say it's three months each season // but it doesn't really make sense // cause this march should be / spring // so you have december // an{d} january an{d} february / are // ul / winter months / but /// then you never know // i think one year they said it was // actually warmer // eh from new year's eve / than it was from midsummer // so you never really know < | >  
@ < VFB; CPUE/A agreement > ... ..  
\$Cm1: < mhm >  
@ < VFB; CPU >, < GFB head: nods; CPU >  
\$Cm1: < **m** >  
@ < **VFB; CPU** >

Unimodal vocal-verbal micro-feedback *ah* and *no* are used to express some additional emotions and attitudes to CPU like surprise and sympathy (see both examples in Excerpt 4).

**Excerpt 4: Unimodal vocal-verbal micro-feedback *ah* and *no*, respectively (from Dial. 3)**

\$Cf1: <1 ul >1 // <2 yeah i would like to >2 / but ul when i searched the izu // university or // university // there's no // ul bachelor course // in english ... ..  
\$Sm1: <1 **ah** >1 / <2 no >2  
@ <1 **VFB; CPUE/A surprise** >1

---

<sup>44</sup> The excerpts presented in this thesis are extracted from the transcriptions of the conversation data studied. More details of the transcription conventions and coding schemes applied can be found in Appendix A and Appendix B. In order to present the examples with good readability, some transcription and coding (e.g., overlapping, vocal-verbal and gestural annotations) that are not directly related to the exemplified micro-feedback but make it difficult to read and understand have been skipped or simplified in the excerpt presentations of this thesis.

<sup>45</sup> The mark "... .." used in this thesis indicates that there are more contexts but they are not presented in the excerpt examples.

@ <2 VFBE; CPUE/A surprise/uncertainty >2, <2 GFBE head: shakes; CPUE/A surprise/uncertainty >2, <2 GFBE eyebrows: rise; CPUE/A surprise/uncertainty >2  
 \$Cf1: < | > about this part / no // ... ...  
 \$Sm1: < **no** >  
 @ < **VF**; **CPUE/A** sympathy >

There are also unimodal vocal-verbal micro-feedback phrases such as *yeah yeah yeah*, *yeah yeah*, *ah yeah*, *yeah okay*, *ah okay*, and *okay okay* that are commonly used by the participants to signal sufficient understanding (see also Table 5.3 above). These micro-feedback expressions most frequently communicate CPU and sometimes express in addition the emotion and attitude of certainty and evaluative opinion agreement (see Excerpts 5 to 7).

**Excerpt 5: Unimodal *yeah yeah yeah* (from Dial. 7)**

\$Sf2: < o:kay > // but you know we swedes // we don't // really pronounce // it's really hard  
 @ < VF; CPU >  
 \$Cm2: <1 **yeah yeah yeah** >1 // i know it's hard for you to pronounce <2 that z >2  
 @ <1 **VF**; **CPUE/A** agreement/certainty >1  
 @ <2 GFB head: Sf2 nods; CPUE/A agreement/certainty >2

**Excerpt 6: Unimodal micro-feedback *okay okay* (from Dial. 8)**

\$Sm2: <1 eh >1 <2 no no >2 // i'm ah // at a: school // eh | // not a elementary school // but a grid li+ ml / um | in america they call high school // we call it gymnasium // okay and ul // it's / a school // at vasa // you know where it i+ / +s situated vasa ... ...  
 \$Cm2: <1 **okay okay** >1 <2 yeah yeah / yeah yeah yeah >2 // (...)  
 @ <1 **VF**; **CPU** >1  
 @ <2 VF; CPU >2, <2 GFB head: tilts; CPU >2

**Excerpt 7: Unimodal *yeah okay* (from Dial. 1)**

\$Cm1: < yes > // and i find a lot of uh // the design studio // in in in / in sweden // yeah there is a website core code uh core seven seven // so is a lot of company / you can find there // and they will / yeah they will / give you a // a connection / to the uh // uh / the job seeking // uh website //  
 \$Sm1: < **yeah okay** >  
 @ < **VF**; **CPU** >  
 \$Cm1: so it's quite ... ...  
 \$Sm1: < ah >  
 @ < VF; CPUE/A interest >

Unimodal vocal-verbal micro-feedback related to misunderstanding will be discussed in the next section.

## In relation to misunderstanding

First, in the data it is found that only the Chinese participants used unimodal vocal-verbal micro-feedback related to misunderstanding and not the Swedish participants.

Table 5.4. The unimodal vocal-verbal feedback expressions that are in relation to misunderstanding (VFB = vocal-verbal micro-feedback).

---

### Unimodal vocal-verbal *eh yeah eh* and *yeah*:

\$Sm1: is < that > in some weeks or / ( is it / i don't know )  
@ < *that* refers to do the thesis project that the interlocutors have been talking about >  
\$Cm1: <1 **eh yeah eh** >1 // <2 **yeah** >2 <3 about ul >3 totally is ul you'll last ul // ul: // fourteen /  
fourteen days // fourteen days // <4 *yeah* >4  
@ <1 **VFB; CPmisUE/A** hesitation >1  
@ <2 **VFB; CPmisUE/A** agreement >2  
@ <3 GFB head: Sm1 nods; CPU >3  
@ <4 VFB; CPU confirmation >4  
\$Sm1: < fourteen days >  
@ < VFBE; CPUE/A uncertainty >, < GFBE; head: up-nod; CPUE/A uncertainty >, < GFBE; eyebrows:  
rise; CPUE/A uncertainty >  
(extracted from Dial. 1)

---

### Unimodal vocal-verbal *Klura* (participant's name):

\$Cm1: i'm < yang fai >  
@ < name: person >, < hand stop: Cm1, Sf1 shake hands >, < general face stop Sf1 smile >  
\$Sf1: i'm < karin >  
@ < name: person >  
\$Cm1: < **klura** >  
@ < mispronounced word: karin >, < **VFB; CmisPmisUE/A** uncertainty >  
\$Sf1: < karin >  
@ < VFB; CPUE/A certainty >  
\$Cm1: karin / oh yeah karin okay / are are you a swe+ swedish ... ...  
(extracted from Dial. 2)

---

As can be seen from Table 5.4, when the Chinese participant says *yeah*, it does not mean “yes I understand you” as it normally does. Instead, it only signals “I want to continue the communication, I have heard what you said, and I am processing it”. In such cases of misunderstanding, the vocal-verbal micro-feedback *yeah* is usually associated with hesitation or uncertainty. How unimodal vocal-verbal micro-feedback is used to signal non-understanding will be addressed next.

## Signalling non-understanding

A couple of the unimodal vocal-verbal micro-feedback expressions are also found to be used to signal non-understanding (see Table 5.5).

Table 5.5. Non-understanding cases signalled by the unimodal vocal-verbal micro-feedback (VFBE = vocal-verbal micro-feedback which has an eliciting function).

---

**Unimodal vocal-verbal micro-feedback *sorry*:**

\$Sf1: ... .. shelter for women ... ..  
 \$Cm1: <1 what's that >1 | <2 **sorry** >2 ... ..  
 @ <1 VFBE; CP-UE/A uncertainty/embarrassment >1, <1 GFB general face: chuckle; CP-UE/A embarrassment >1  
 @ <2 **VFBE; CP-UE/A** uncertainty/embarrassment >2  
 \$Sf1: <1 yeah >1 <2 um >2 // i can understand you an{d} i // i didn't know they existed before ... ..  
 (extracted from Dial. 2)

---

**Unimodal micro-feedback *eh what do you mean*:**

\$Sf2: are you from a city or  
 \$Cm2: <1 **eh / what do you mean** >1 // <2 city or / countryside eh (...) >2  
 @ <1 **VFBE; CP-UE/A** uncertainty >1  
 @ <2 VFBE; CP-UE/A uncertainty >2, <2 GFB general face: smile; CP-UE/A embarrassment >2, <2  
 GFBE hand: hands move from one side to another to show symbolic distance between city and  
 countryside; CP-UE/A uncertainty >2  
 \$Sf2: <1 no >1 / are you from ci+ / <2 yeah city or countryside >2 or wherever  
 @ <1 VFB; CPmisUE/A disconfirmation >1, <1 GFB gaze: down; CPmisUE/A uncertainty >1  
 @ <2 VFB; CPUE/A confirmation >2  
 (Extracted from Dial. 7)

---

As presented in the excerpts above, the participants used the unimodal vocal-verbal micro-feedback *sorry* and *what do you mean* to express that they did not understand the information communicated. While signalling non-understanding, the participants also expressed uncertainty and tried to elicit further information from the other interlocutor.

## 5.2.2 Unimodal gestural micro-feedback

After having investigated how unimodal vocal-verbal micro-feedback are related to sufficient understanding, misunderstanding, and non-understanding, I will study unimodal gestural micro-feedback in this section.

Table 5.6 presents an overview of unimodal gestural micro-feedback in relation to sufficient understanding, misunderstanding, and non-understanding in the data studied. Most of the unimodal gestural micro-feedback expressions made by the participants were used to signal sufficient understanding (see Table 5.6). There are only two cases of non-understanding in the research data that were used by one Swede and one Chinese. Neither the Swedes nor the Chinese had misunderstanding related to any unimodal gestural micro-feedback.

Table 5.6. Overview of the unimodal gestural micro-feedback related to sufficient understanding, misunderstanding, and non-understanding.

Micro-feedback	Participant	Sufficient understanding	Mis-understanding	Non-understanding	Total
Unimodal gestural	Swedish	177	0	1	178
	Chinese	164	0	1	165
	Total	341	0	2	343

### For sufficient understanding

First, sufficient understanding is investigated. As presented in Table 5.7, the participants used nods, nod, smile, up-nod, up-nods, head shakes, head tilt, and eyebrow rise as the most common unimodal gestural micro-feedback expressions to communicate sufficient understanding.

Examples of how the unimodal gestural micro-feedback nods, nod, smile, up-nod, up-nods, head shakes, and eyebrow rise were used to signal sufficient understanding are presented in Excerpts 8 to 12.

Table 5.7. The most common unimodal gestural micro-feedback (GFB) that signals sufficient understanding (F. = frequency).

Unimodal GFB	F.	/1,000 words	/min.
nods	206	20.34	3.16
nod	32	3.17	0.5
smile	27	2.68	0.42
up-nod	19	1.89	0.29
up-nods	13	1.3	0.2
head shakes	6	0.58	0.09
head tilt	6	0.58	0.09
eyebrows rise	4	0.38	0.06
head forward	2	0.19	0.03
head complex	2	0.19	0.03
hand move	2	0.19	0.03
Others (F. = 1)	22	2.18	0.34
Total	341	33.67	5.24

### **Excerpt 8: Unimodal gestural micro-feedback nods (from Dial 1)**

\$Cm1: or mobile / what's <1 the color and uh >1 // and uh we also need to / <2 analyse >2 // about uh which product could be // more acceptable // by more users // something like that

@ <1 **GFB head: Sm1 nods; CPU** >1

@ <2 **GFB head: Sm1 nods; CPU** >2

### **Excerpt 9: Unimodal micro-feedback smile (from Dial. 3)**

\$Sm1: <1 yeah >1 in the same class / el+ as eh eh <2 jia >2

@ <1 VFB; CPUE/A confirmation >1

@ <2 name: person >2

\$Cf1: <1 lu jia >1 / <2 okay >2

@ <1 VFB; CPU >1, <1 GFB head: nods; CPU >1, <1 name: person >1

@ <2 VFB; CPU >2

\$Sm1: <1 lu jia <2 lu jia hm >2 >1 so are / i'm also doing a research but not like this ... ..

@ <1 VFBE; CPUE/A checking correct up-take repetition >1, <1 GFBE head: nods; CPUE/A checking correct up-take repetition >1

@ <2 **GFB general face: Cf1 smile; CPUE/A amusement** >2 ... ..

### **Excerpt 10: Unimodal up-nod (from Dial. 3)**

\$Sm1: <1 <2 oh >2 right >1 // that is bad // so many people are ul queuing for // is a / oh // they should have more {uni}versi{ty} teachers and (...) | m+ m+ / more classes than

@ <1 VFB; CPU >1

@ <2 **GFB head: Cf1 up-nod; CPUE/A agreement** >2 ... ..

### **Excerpt 11: Unimodal head shakes (from Dial. 4)**

\$Sf1: maybe it wouldn't care as much <1 like >1 // i don't know if it is ba+ i think it's // yeah i think sw+ sweden / we talked about it before some people that // mo+ most people when you / graduate then you // work for a while / you travel // you do other things and then maybe you go to university / and it's just very // relaxed ... ..

@ <1 **GFB head: Cf1 shakes; CPUE/A agreement** >1 ... ..

### **Excerpt 12: Unimodal micro-feedback eyebrow rise and unimodal nod (from Dial. 4)**

\$Sf1: i think that // it seems lite it's swedish people // they always tend to choose what they think is funny // and what they think interesting // and they don't really think about / if it can actually work // but it's // <8 eh >8 and i think that // many times <9 / you will >9 <10 have >10 // very / like skilled people <11 from other >11 countries coming to sweden and studying something // and // they become / very good // and they learn very good professions / and swedish people tend to take / these // like / courses there just yeah this is fun and // and then we are unemployed ... ..

@ <8 GFB head: Cf1 nods; CPU >8

@ <9 GFB head: Cf1 nods; CPU >9

@ <10 **GFB eyebrows: Cf1 rise; CPUE/A interest** >10

@ <11 **GFB Cf1 nod; CPU** >11 ... ..

As presented in the excerpts above, these unimodal gestural micro-feedback expressions that were used to signal sufficient understanding are also used to communicate the emotions and attitudes of, for example, amusement and interest, as well as the evaluative opinion of agreement.

## Signalling non-understanding

Since no misunderstanding is communicated through unimodal gestural micro-feedback, non-understanding will be studied in this section.

Table 5.8. Unimodal gestural micro-feedback (GFB) signals non-understanding.

---

### Unimodal eyebrow rise:

\$Sm1: how did you end up here < | >  
@ < chuckle >  
\$Cf1: < mhm >  
@ < VFB; C-P-UE/A uncertainty/politeness >, < GFB head: nod; C-P-UE/A uncertainty/politeness >  
\$Sm1: how did you < | > end up here with the // in this // research  
@ < **GFB eyebrows: rise; C-P-UE/A uncertainty** >  
\$Cf1: <1 yeah >1 u:h // like / a a friend <2 of // >2 julia  
@ <1 VFB; CPU >1  
@ <2 GFB head: Sm1 nods; CPU >2  
(extracted from Dial. 3)

---

### Unimodal gaze at: (*Note. "Gaze" at here is more of a long lasting "stare at" and is different from the normal "look at" in the daily life.*)

\$Sf2: s:+ still // yeah / yeah // but as we used to say // eh i mean city one is a big small city // because  
\$Cm2: <1 what >1 <2 | >2  
@ <1 VFBE; CP-UE/A uncertainty/interest >1  
@ <2 **GFB gaze: at; CP-UE/A uncertainty/interest** >2  
\$Sf2: it's eh // i mean we here in sweden we say this is a big um // this is the l+ description of city // so  
it's a yeah it's a bi+ big town // but it's really not ml ml // i mean it's really small town / so we say  
/ it's a big // small town ... ..  
\$Cm2: < yeah > i heard about it ... ..  
(extracted from Dial. 7)

---

In the data, the participants used the unimodal gestural micro-feedback eyebrow rise and gaze at to signal non-understanding (see excerpts in Table 5.8). As presented in the excerpts, the participants usually express uncertainty (about what has been perceived) and interest (in knowing more) through unimodal gestural micro-feedback, that is, in relation to non-understanding.

## 5.3 Multimodal micro-feedback

In this section, how multimodal micro-feedback is used in relation to sufficient understanding, misunderstanding, and non-understanding, respectively, will be investigated.

As presented in Table 5.9, most of the multimodal micro-feedback expressions in the data were used to signal that the interlocutors have sufficiently understood one another (with a frequency of 579). The participants used more multimodal micro-

feedback expressions to convey non-understanding (with an occurrence of 19) than misunderstanding (with an occurrence of 6).

Table 5.9. *Multimodal*<sup>46</sup> micro-feedback related to the different types of understandings.

Micro-feedback	Participant	Sufficient understanding	Mis-understanding	Non-understanding	Total
Multimodal	Swedish	340	5	9	354
	Chinese	239	1	10	250
	Total	579	6	19	604

Regarding the cultural differences shown in this particular data, the Swedes show higher expressivity when using multimodal micro-feedback to convey sufficient understanding and misunderstanding than the Chinese, 340 compared to 239 and 5 to 1 (see Table 5.9). The data show that the Swedish and the Chinese participants used roughly the same number of multimodal micro-feedback expressions to signal non-understanding, with a frequency of 9 and 10, respectively.

### 5.3.1 For sufficient understanding

The participants used *yeah+nods*, *chuckle*, *yeah+nod*, *m+nods*, *laughter*, *okay+nods*, *mhm+nod*, *okay+up-nod*, *yeah+up-nod*, and *okay+nod* (with a frequency of no less than 10) as the most common multimodal micro-feedback to signal sufficient understanding (see Table 5.10 and Excerpts 13 to 16).

---

<sup>46</sup> Laughter is regarded as one multimodal unit, consisting of sound and facial gesture. Chuckle is also treated as a multimodal unit.

Table 5.10. The most frequent multimodal micro-feedback, which is used to signal sufficient understanding, shown with the vocal-verbal and gestural components (F. = frequency, F. ≤ 2 means the frequency is 2 or below).

Vocal-verbal part	Gestural part	F.	/1,000 words	/min.
<i>yeah</i>	nods	62	6.11	0.92
–	chuckle	44	4.35	0.67
<i>yeah</i>	nod	31	3.06	0.48
<i>m</i>	nods	28	2.76	0.41
–	laughter	16	1.58	0.24
<i>okay</i>	nods	12	1.19	0.19
<i>mhm</i>	nod	10	0.99	0.16
<i>okay</i>	up-nod	10	0.99	0.16
<i>yeah</i>	up-nod	10	0.99	0.16
<i>okay</i>	nod	10	0.99	0.16
<i>m</i>	up-nods	9	0.89	0.14
<i>yeah</i>	up-nods	8	0.79	0.12
<i>m</i>	nod	8	0.78	0.12
<i>yes</i>	nod	8	0.79	0.13
<i>m</i>	up-nod	6	0.59	0.09
<i>mhm</i>	nods	6	0.59	0.09
<i>ah</i>	up-nod	5	0.49	0.08
<i>yeah</i>	smile	4	0.39	0.06
<i>uhu</i>	nods	4	0.39	0.06
<i>yes</i>	nods	4	0.39	0.06
<i>aha</i>	nods	3	0.3	0.05
<i>oh</i>	nods	3	0.3	0.05
–	giggle	3	0.3	0.05
<i>ah okay</i>	up-nod	3	0.3	0.05
<i>ah okay</i>	up-nods	3	0.3	0.05
<i>yeah yeah</i>	nods	3	0.3	0.05
<i>okay</i>	up-nods	3	0.3	0.05
<i>yeah</i>	gaze sideways	3	0.3	0.05
<i>yeah</i>	chuckle	3	0.3	0.05
<i>yeah</i>	smile+nods	3	0.3	0.05
<i>yeah / okay</i>	up-nods	3	0.3	0.05
<i>yeah okay</i>	up-nods	3	0.3	0.05
Others (F. ≤ 2)		248	24.47	3.79
Total		579	57.17	8.89

### Excerpt 13: The example of *yeah*+nods (from Dial. 3)

\$Sm1: how did you < | > end up here with the // in this // research ... ..  
\$Cf1: <1 yeah >1 u:h // like / a a friend <2 of // >2 julia  
@ <1 VFB; CPU >1  
@ <2 GFB head: Sm1 nods; CPU >2  
\$Sm1: < yeah >  
@ < VFB; CPU >, < GFB head: nods; CPU >

### Excerpt 14: Example of *yeah*+nod (from Dial. 5)

\$Sm2: we call it peking  
\$Cf2: <1 | >1 <2 yeah >2 peking // en // and ul  
@ <1 GFB general face: chuckle; CPUE/A friendliness >1  
@ <2 VFB; CPU confirmation >2, <2 GFB head: nod; CPUE/A confirmation >2 ... ..

### Excerpt 15: Example of chuckle signalling sufficient understanding (from Dial. 7)

\$Cm2: i like the weather here  
\$Sf2: <1 | >1 <2 okay >2 <3 s+ (...) >3  
@ <1 GFB general face: chuckle; CPUE/A amusement/surprise >1  
@ <2 VFB; CPUE/A surprise/amusement >2  
@ <3 general face: chuckling >3  
\$Cm2: e+ // except winter  
\$Sf2: < yeah > it's so // funny i'm so surprised that you say that // because i've got she said the same  
thing that // that she liked the weather  
@ < VFB; CPU >  
\$Cm2: <1 yeah >1 that's the weather // ml <2 but only summer >2  
@ <1 VFB; CPU >1, <1 GFB gaze: down; CPUE/A hesitation >1  
@ <2 GFB general face: Sf2 chuckle; CPUE/A amusement >2

### Excerpt 16: Example of *m*+nod (from Dial. 5)

\$Sm2: <1 hm >1 // maybe want to get some eh // experience // of the other world and <2 other cultures  
and then >2  
@ <1 VFB; CmisPmisUE/A hesitation/consider >1, <1 GFB gaze: sideways; CmisPmisUE/A hesitation/con-  
sider >1  
@ <2 GFB head: Cf2 nods; CPU >2  
\$Cf2: < m >  
@ < VFB; CPU >, < GFB head: nod; CPU >

### Excerpt 17: Examples of Swedish up-nod and up-nods in the multimodal micro-feedback which signals sufficient understanding (extracted from Dial. 7)

\$Cm2: eh // i'm ah // s+ second year master student in chalmers ... ..  
\$Sf2: < m: >  
@ < VFB; CPU >, < GFB head: up-nod; CPU > ... ..  
\$Sf2: < | > do you like it  
\$Cm2: eh sort of  
\$Sf2: < m >  
@ < VFB; CPU >, < GFB head: up-nods; CPU >, < GFB general face: Cm2 smile; CPUE/A awkwardness >  
\$Cm2: but ul sometimes it's // boring ... ..

The data shows that the Swedes tend to use more up-nod(s) than the Chinese as the gestural component of the multimodal micro-feedback that signals sufficient understanding (see Table 5.10 and Excerpt 17).

### 5.3.2 In relation to misunderstanding

There are six instances of misunderstanding that are related to multimodal micro-feedback expressions in the data studied.

*Table 5.11.* Multimodal micro-feedback related to misunderstanding (VFB = unimodal vocal-verbal micro-feedback, GFB = unimodal gestural micro-feedback).

---

**Multimodal micro-feedback of *lu chang* (participant's name) combined with smile+nod:**

\$Cf1: < | > hi // hello // you can call me huang tian ... ...  
 \$\$f1: < **lu chang** >  
 @ < **GFB general face: smile; CPmisUE/A** friendliness >, < mispronounced word >, < name: person >,  
 < **VFBE; CmisPmisU** uncertainty >, < **GFBE head: nod; CmisPmisUE/A** uncertainty >  
 \$Cf1: < huang tian >  
 @ < **GFB head: nod; CPU** correction >, < **VFB; CPU** correction >  
 (Extracted from Dial. 4)

---

**Multimodal micro-feedback of *hm* together with gaze sideways:** (*Note.* Gaze at here is more of a long lasting stare at and is different from the normal look at in the daily life.)

\$Cf2: because i / i think / ul i mean | how can you guess i sturdied in university ... ...  
 \$Sm2: <1 hm >1 // maybe want to get some eh // experience // of the other world and <2 other cultures and then >2  
 @ <1 **VFB; CmisPmisUE/A** hesitation/consider >1, <1 **GFB gaze: sideways; CmisPmisUE/A** hesitation/consider >1  
 @ <2 **GFB head: Cf2** nods; CPU >2  
 \$Cf2: < m >  
 @ < **VFB; CPU** >, < **GFB head: nod; CPU** >  
 (Extracted from Dial. 5)

---

**Multimodal micro-feedback of *no* together with gaze down:**

\$\$f2: are you from a city or  
 \$Cm2: <1 eh / what do you mean >1 // <2 city or / countryside eh (...) >2  
 @ <1 **VFBE; CP-UE/A** uncertainty >1  
 @ <2 **VFBE; CP-UE/A** uncertainty >2, <2 **GFB general face: smile; CP-UE/A** embarrassment >2, <2  
**GFBE hand: hands** move from one side to another to show symbolic distance between city and  
 countryside; CP-UE/A uncertainty >2  
 \$Sf2: <1 **no** >1 / are you from ci+ / <2 yeah city or countryside >2 or wherever  
 @ <1 **VFB; CPmisUE/A** disconfirmation >1, <1 **GFB gaze: down; CPmisUE/A** uncertainty >1  
 @ <2 **VFB; CPUE/A** confirmation >2  
 (Extracted from Dial. 7)

---

---

**Multimodal micro-feedback of *no I don't drive I don't drive* combined with hands movement to show symbolic meaning of *no*:**

\$Cm2: < okay > // you should eh // elder than th+ / eighteen i think  
@ < VFB; CPU >, < GFB gaze: down; CPUE/A hesitation >  
\$Sm2: < huh >  
@ < VFBE; CP-UE/A uncertainty >, < GFBE eyebrows: frown; CP-UE/A uncertainty >  
\$Cm2: eighteen // i mean // eh the age allow to // drive ... ..  
\$Sm2: <1 drive >1 <2 **no i i don't drive i don't drive** >2 // i'm taking {the} bus  
@ <1 VFB; CPUE/A surprise >1, <1 GFB eyebrows: rise; CPUE/A surprise >1  
@ <2 **VFB; CPmisU** >2, <2 **GFB hand: hands move to show symbolic meaning of no; CPmisU** >2  
\$Cm2: < no no > i (...) i mean the the the a+ age  
@ < VFB; CPU correction >  
(Extracted from Dial. 8)

---

**Two multimodal micro-feedback examples of *one year* combined with nod and *yeah* together with nod:**

\$Sf2: < m > / how long have you been here  
@ < VFB; CPU >  
\$Cf2: ml / half and f+ one year  
\$Sf2: < **one year** >  
@ < **VFBE; CmisPmisUE/A** uncertainty >, < GFBE head: nod; **CmisPmisUE/A** uncertainty >  
\$Cf2: < **yeah** >  
@ < **VFB; CPmisUE/A** confirmation >, < **GFB head: nod; CPmisUE/A** confirmation >  
\$Sf2: <1 | >1 <2 okay >2 // <3 m >3 how do you like it then  
@ <1 click >1  
@ <2 VFB; CPU >2, <2 GFB head: up-nods; CPU >2  
@ <3 VFB; CPU >3  
(Extracted from Dial. 6)

---

As can be seen from the excerpts above, half of the time the multimodal micro-feedback expressions that are related to misunderstanding consist of a repetition of the perceived message, which is normally a noun phrase or a word *yeah*, and an assertive gesture nod for information confirmation. Also, misunderstanding may not be detected by the interlocutors, and it can result in further misunderstandings in the interaction. As in the Chinese case of misunderstanding, Sf2 asks Cf2 *how long have you been here* and Cf2 says *half and one year*. Then, Sf2 confirms this by repeating the misperceived and misunderstood information *one year*. In response, Cf2 says *yeah*, which shows that she has also misunderstood Sf2's misunderstood information.

### 5.3.3 Signalling non-understanding

Regarding non-understanding, in the studied data the participants demonstrated 19 cases in total, which were signalled by means of multimodal micro-feedback expressions (see Table 5.12).

As can be seen from the excerpts below, the participants used multimodal micro-feedback, which often consisted of vocal-verbal expressions *what*, *sorry*, and *huh* and gestural expressions eyebrow rise, eyebrow frown, and head forward to signal non-understanding. Uncertainty and hesitation are typically expressed. Besides these, the participants had communicative gaze movements such as gaze sideways and gaze at (which express more attention, thoughtfulness, and eagerness to know more) and facial expressions such as chuckle and smile (which sometimes express politeness and embarrassment) as components of the multimodal micro-feedback, which also signals non-understanding.

These vocal-verbal and gestural components of multimodal micro-feedback are very often used as eliciting devices to express that the interlocutor has not understood the other interlocutor's communicated information and is willing to know more about it in order to develop further understandings (see Table 5.12).

Table 5.12. Multimodal micro-feedback signals non-understanding.

---

**Multimodal micro-feedback of *förlåt* combined with head forward:**

\$Cf1: < yes > // and it's the first one i thi+ // an{d} it's first of all // when we write names // karin first  
 ... ..  
 \$Sf1: < **fö(r)låt** >  
 @ < **GFBE head: forward; C-P-UE/A uncertainty** >, < other language: swedish >, < **VFBE; C-P-UE/A uncertainty** >  
 (Extracted from Dial. 4)

---

**Multimodal micro-feedback of *huh* together with eyebrow rise+ head forward+ mouth open:**

\$Cf2: ... .. you guess // i sturdied there ... ..  
 \$Sm2: < **huh** >  
 @ < **VFB; CP-UE/A uncertainty** >, < **GFBE eyebrows start: rise; CP-UE/A uncertainty** >, < **GFBE head: forward; CP-UE/A uncertainty** >, < **GFB mouth: open; CP-UE/A uncertainty** >  
 \$Cf2: because i / i think / ul i mean | how can you guess i sturdied in university ... ..  
 \$Sm2: <1 hm >1 // maybe want to get some eh // experience // of the other world and <2 other cultures and then >2  
 @ <1 VFB; CmisPmisUE/A hesitation/consider >1, <1 GFB gaze: sideways; CmisPmisUE/A hesitation/consider >1  
 @ <2 GFB head: Cf2 nods; CPU >2  
 (Extracted from Dial. 5)

---

**Multimodal micro-feedback example of chuckle:**

\$Cf2: here // <1 okay >1 / <2 yeah okay >2 <3 ha det bra >3 ... ..  
 @ <3 other language: swedish >3  
 \$Sm2: so <1 /// >1 but ul // when you here s:weden ... ..  
 @ <1 **GFB general face: Cf2 chuckle; CP-UE/A politeness/friendliness** >1 ... ..  
 (Extracted from Dial. 5)

---

---

**Multimodal micro-feedback of *what* combined with head forward+posture forward:**

\$Cf2: ( at least you're ) with your boyfriend ... ...  
\$Sf2: < yeah >  
@ < VFB; CPU >  
\$Cf2: you have beautiful eyes // eyes // you have beautiful eyes  
\$Sf2: < **what** >  
@ < **VFBE; C-P-UE/A** uncertainty >, < **GFBE head: forward; C-P-UE/A** uncertainty >, < **GFBE posture: forward; C-P-UE/A** uncertainty >  
\$Sf2: < oh > thank you // it's very kind ... ...  
(Extracted from Dial. 6)

---

**Multimodal micro-feedback of *sorry* combined with eyebrow rise+gaze at:**

\$Sf2: i'm sandra  
\$Cf2: eh my name is eh li yun // li yun  
\$Sf2: < okay >  
\$Cf2: m: eh / i co{me} from china //  
\$Sf2: it's // similar  
\$Cf2: < **sorry** >  
@ < **VFBE; C-P-UE/A** uncertainty >, < **GFBE eyebrows: rise; C-P-UE/A** uncertainty >, < **GFBE gaze: at; C-P-UE/A** interest >  
\$Sf2: < li yun an{d} sandra / it's similar in a way >  
@ < comment: the original names before the anonymisation were a bit similar >  
(Extracted from Dial. 6)

---

**Multimodal micro-feedback of *mhm* together with nod:**

\$Sm1: how did you end up here < | >  
@ < chuckle >  
\$Cf1: < **mhm** >  
@ < **VFB; C-P-UE/A** uncertainty/politeness >, < **GFBE head: nod; C-P-UE/A** uncertainty/politeness >  
\$Sm1: how did you < | > end up here with the // in this // research  
(Extracted from Dial. 3)

---

Discussions regarding micro-feedback in relation to sufficient understanding, misunderstanding, and non-understanding will be presented as in the following section.

## 5.4 Discussion

The present analysis shows that communicative micro-feedback expressions mainly signal sufficient understanding (with a frequency of 1,256) and, comparatively, are only to a small extent related to misunderstanding (9 cases) and non-understanding (23 occurrences) (see Table 5.1). This suggests that the participants can understand each other fairly well in their intercultural conversations, and that misunderstanding is least signalled or conveyed through micro-feedback expressions. The unimodal and multimodal micro-feedback expressions are similar in size, with the former being slightly larger. Within the unimodal micro-feedback expressions, the numbers of

vocal-verbal and gestural expressions are roughly the same. Furthermore, multimodal micro-feedback expressions are more frequently used than unimodal ones in relation to misunderstanding and non-understanding. For sufficient understanding, the five most frequent unimodal vocal-verbal micro-feedback expressions are *yeah*, *okay*, *m*, *ah*, and *yes*, and the most common unimodal gestural ones are nods and smile. Also, the top five multimodal ones are *yeah+nods*, *chuckle*, *yeah+nod*, *m+nods*, and *laughter*. Regarding misunderstanding, there are three unimodal vocal-verbal micro-feedback expressions in use. They are *eh yeah eh*, *yeah*, and a participant's name, which are usually expressed with hesitation or uncertainty. None of the misunderstanding cases is expressed through unimodal gestural micro-feedback. Further, half the multimodal micro-feedback expressions consist of nod together with *yeah* or a noun phrase (repetition of the perceived information). The unimodal micro-feedback expressions *sorry*, *what do you mean*, *eyebrow rise*, and *gaze at* are used to signal non-understanding. The most common multimodal micro-feedback expressions include head forward or eyebrow rise together with *sorry*, *what*, or *huh*. With micro-feedback signalling non-understanding, the interlocutors express uncertainty and hesitation and try to elicit further information.

These findings will be interpreted by first looking at the idea that misunderstanding is ubiquitous in conversation. Then, the unimodal gestural micro-feedback nods and up-nods will be discussed. Next, cases of misunderstanding conveyed by the vocal-verbal expression *yeah* and gestural head nod will be presented. Following that, I will turn to multimodal micro-feedback with a focus on its vocal-verbal and gestural composition. Lastly, some implications from the present analysis for communication technology applications will be discussed.

#### 5.4.1 The ubiquity of misunderstanding in conversation

The view that misunderstanding is ubiquitous in human conversation, presented by, for example, Fraser (1993) and Dascal (1999), is not reflected in the present analysis of micro-feedback expressions in first encounters. Only around 2.4% (see Table 5.1) of the micro-feedback expressions are related to misunderstanding and non-understanding, with occurrences of 0.7% and 1.7%, respectively.

There are some possible explanations. As has been found in the data, nod and nods are the most frequent unimodal gestural micro-feedback expressions and they are typically used to signal sufficient understanding. According to Ishi, Ishiguro, and Hagita (2014), the frequency of nod and nods increases in interpersonal communication when the interpersonal relationship is distant. Thus, it suggests that the distant interpersonal relationship, as in the present data on first encounters, not only in-

creases the use of nod(s) in conversation, but also the frequency of sufficient understanding over misunderstanding and non-understanding. Also, it is possible that the interlocutors wish to be polite and not indicate many understanding problems such as misunderstanding and non-understanding because of the social situation, especially in such first encounter conversations. Another explanation could be that although people have more difficulties in understanding in intercultural communications than mono-cultural ones, they make an effort to compensate for the risk of misunderstanding by utilising practices in sense-making and information sharing. For instance, Kaur (2010) found that the speakers of world English used various confirmation and clarification procedures, for example, repetition and paraphrase to a larger extent than did the native English speakers, in order to enhance mutual understanding. Thus, the number of misunderstanding and non-understanding cases, especially the misunderstanding ones, appears to be low in this study.

Besides these activity type explanations, the coding scheme of understanding used in the present analysis may vary to some extent from other studies, for example, in terms of what is classified as misunderstanding. Earlier studies, as discussions in literature reviews and theory chapters, have primarily focused on qualitative analyses of understanding in conversation rather than quantitative ones. They do not seem to count understanding or misunderstanding in the same way as this study. Also, perhaps micro-feedback has a communicative function of signalling shallow understandings, and it is restricted when signalling deeper levels of understandings like understanding problems. Thus, this requires further research.

#### 5.4.2 Unimodal gesture that exclusively signals sufficient understanding

The most frequent unimodal gestural micro-feedback found in the data is head nod and nods. This result corresponds well with other findings in related studies of communicative *feedback* (i.e., micro-feedback) in several languages, such as Swedish and Finnish (Navarretta, Ahlsén, Allwood, Jokinen, & Paggio, 2012), Danish (Paggio & Navarretta, 2013), and Japanese (Ishi et al., 2014).

Up-nod and up-nods are the second most frequent unimodal gestural micro-feedback to signal sufficient understanding, with a frequency of 32 (see Table 5.7). A high frequency can be also seen in the multimodal data, in which up-nod and up-nods constitute the gestural part of the micro-feedback in over 60 cases. This is probably because the Swedes frequently use up-nods to communicate CPU in interaction (see Navarretta et al., 2012) and the Chinese are very likely influenced by the Swedes through the process of co-activation (similarly named as *mimicry* in Goffman, 1974; *matching* in Couper-Kuhlen, 1996; *co-activation* in Allwood & Lu, 2011). This co-

activation effect is substantial, in that the Chinese almost never use up-nod and up-nods in their mono-cultural conversations (Lu, 2012), but in the Chinese–Swedish data studied, the Chinese and the Swedes use up-nod and up-nods with similar frequencies of 13 and 19, respectively (see Table 5.7 in Section 5.2.2). However, as suggested in the data focused on, there is a large difference between the Chinese and the Swedish participants when using up-nod and up-nods as a component of the multimodal micro-feedback; the Chinese use only one tenth of that used by the Swedes. It suggests that the interlocutors are more easily influenced in their unimodal gestural behaviours than multimodal ones, which is probably due to the higher complexity of multimodal behaviours. Thus, the high frequency of nod, nods, up-nod, and up-nods signalling sufficient understanding in the analysis can possibly be explained by the nodding conventions of the Chinese and the Swedes and the interactional co-activation processes.

Furthermore, it is found in this study that unimodal gestural micro-feedback almost exclusively signals sufficient understanding. No cases of misunderstanding and only two cases of non-understanding (eyebrow rise and gaze at) are conveyed by unimodal gestural micro-feedback. All the unimodal head movements (particularly head nod and nods) are found to exclusively signal sufficient understanding rather than misunderstanding or non-understanding.

### 5.4.3 *Yeah* and nod related to misunderstanding

The analysis has also found that the micro-feedback *yeah* and nod can be used to express misunderstanding. The empirical data (see the first example presented in Table 5.4 of Section 5.2.1) show that when a participant says *yeah*, it does not always mean “yes I understand you”. Sometimes, it is related to misunderstanding, which is usually associated with hesitation that is very often communicated through prosodic features. Similarly, the multimodal micro-feedback *yeah*+nod is sometimes also used in relation to misunderstanding (see the third example presented in Table 5.11 of Section 5.3.2).

The multimodal micro-feedback expressions related to misunderstanding sometimes consist of a repetition of the perceived vocal-verbal message, which is usually a noun phrase and an assertive gesture nod for information confirmation (see Table 5.11 also). Very likely, such a misunderstanding can result in further misunderstandings between the interlocutors. As in the third example presented in Table 5.11, interlocutor Sf2 asks Cf2 *how long have you been here* and Cf2 says *half and one year*. Then, Sf2 confirms this by repeating the misperceived and misunderstood information *one year*. In response, Cf2 says *yeah*, which shows that she too has misunder-

stood Sf2's misunderstood information. This example shows that both the interlocutors can sometimes continue communicating without any awareness or correction of the earlier misunderstood information. This point is in line with Weigand's (1999) and Verdonik's (2010) claims that a misunderstanding is not always recognised and corrected by the interlocutors.

#### 5.4.4 Multimodal micro-feedback

As found in the data, around half (based on Table 5.1) of the micro-feedback expressions related to the three types of understandings are multimodal. The relation<sup>47</sup> between the vocal-verbal and gestural components of the multimodal micro-feedback is further investigated. In the data, it is found that most of the relations are congruent, for example, *yeah* and head nod, and very few are conflictual (with 4 occurrences in all), for example, *yeah* and head shake. One case is the vocal-verbal micro-feedback *no* together with the gestural head nods. Another case is *okay* accompanied by a head shake. Also, there are two instances of *no* combined with up-nods (up-nod means "yes" in Swedish). Furthermore, it is clear from the context that all these multimodal micro-feedback expressions that have a conflictual relation (between the vocal-verbal and the gestural components) signal sufficient understanding. No such instances exist with respect to misunderstanding and non-understanding. To my knowledge, no earlier study has investigated the relation between the vocal-verbal and the gestural components of multimodal micro-feedback in relation to understanding. The most closely related study is by Hindmarsh et al. (2011, p. 501), in which they identified "mixed messages" between the verbal and bodily behaviours that occurred simultaneously and jointly expressed understanding. In their study, when a teacher asked a student whether he understood the teacher's instruction, the student gave an affirmative answer verbally although he made numerous efforts in his bodily movements (e.g., shifting his body trunk) in order to see more clearly and obtain further understanding of what the teacher had instructed. In this case, Hindmarsh et al. claim that the verbal behaviour expresses understanding, whereas the bodily action reveals a lack of understanding. Hindmarsh et al.'s "mixed messages" may seem close to the conflictual relations in the present data. However, Hindmarsh et al. studied the conflictual understandings expressed through the verbal and bodily communicative behaviours in general rather than the conflictual relation between the vocal-verbal and

---

<sup>47</sup> The relation between the vocal-verbal and gestural components of multimodal micro-feedback can be congruent, conflictual, or otherwise (e.g., when they neither reinforce nor contradict each other).

gestural components of the particular communicative phenomenon of micro-feedback. Further research is needed in order to acquire better insights into this issue in relation to understanding.

Another interesting finding in the data is that multimodal micro-feedback expressions are more frequently used than unimodal ones in relation to misunderstanding and non-understanding. This is probably because multimodal expressions convey more information (per time and word unit) than unimodal ones. When there is a problem in understanding, multimodal micro-feedback expressions may, thus, repair the interaction more efficiently.

#### 5.4.5 The role of visual modality in signalling understanding

How important is visual modality in signalling understanding through micro-feedback, and what does this imply for communication technology applications? The present analysis has found that visual modality (i.e., gesture) plays an important role in communicating micro-feedback and signalling understanding; here, gesture is involved in around 74% (based on the figures in Table 5.1) of all the micro-feedback expressions that are related to different types of understandings. In addition to this, these gestures are found to be almost entirely limited to the head region in the form of head movements and facial expressions. Hand and posture movements rarely occur. As discussed earlier in Section 5.4.2, among all the micro-feedback head movements, the unimodal ones exclusively signal sufficient understanding. These findings can serve as a basis for developing some guidelines for the design of communication technology applications. Such a design should include visual modality since a large portion of micro-feedback occurs there in relation to understanding in communication. Further, the visual parts of the system, such as the graphical display and motion capture, can be limited to the head region of the participants without compromising the signalling or communication of understanding.

### 5.5 Conclusion of Chapter 5

The aim of this analysis was to explore the modalities of micro-feedback that are related to different types of understandings in eight audio- and video-recorded Swedish–Chinese intercultural communication dialogues in the English language. The unimodal vocal-verbal, unimodal gestural, and multimodal micro-feedback as well as its relation to sufficient understanding, misunderstanding, and non-understanding have been investigated. Two research questions have been investigated. First, how are the auditory and visual modalities involved in micro-feedback expressions that are related to sufficient understanding, misunderstanding, and non-understanding?

Second, what are the typical unimodal and multimodal micro-feedback expressions that signal sufficient understanding, misunderstanding, and non-understanding?

First of all, in the data focused on, it has been found that most of the micro-feedback expressions signal sufficient understanding; a few convey non-understanding, and fewer still are in relation to misunderstanding. This suggests that compared to sufficient understanding and non-understanding, misunderstanding is least signalled through micro-feedback and is most difficult to observe in spontaneous communication. The view that misunderstanding is ubiquitous in conversation (see Fraser, 1993; Dascal, 1999) is not supported by the present analysis of micro-feedback. Further, the results show that sufficient understanding is signalled more through unimodal micro-feedback than multimodal, with roughly the same numbers of unimodal vocal-verbal and unimodal gestural micro-feedback expressions. Misunderstanding involves more multimodal micro-feedback than unimodal vocal-verbal, and it is not signalled through unimodal gestural micro-feedback at all. Besides this, non-understanding is mostly shown by means of multimodal micro-feedback expressions and rarely by unimodal expressions.

When signalling sufficient understanding, the data suggests that the most common unimodal vocal-verbal micro-feedback expressions are *yeah, okay, m, ah, yes, yeah yeah yeah, yeah yeah, ah yeah, yeah okay, ah okay, and okay okay*. Among them, *yeah, okay, m, ah, and yes* are comparable with the five most frequent *backchannel* expressions found in other corpora and studies (Jurafsky et al., 1998; Ward & Tsukahara, 2000). These vocal-verbal micro-feedback expressions are always used to communicate “I hear you, I perceive and understand what you have said, and I would like to continue the conversation with you” (CPU), and sometimes to express emotions and attitudes, for instance, of agreement, disagreement, certainty, amusement, interest, and surprise. The most common gestural micro-feedback expressions are single and repeated head nods, which corresponds well with other studies of communicative *feedback* (i.e., micro-feedback) in several languages, such as Swedish and Finnish (Navarretta et al., 2012), Danish (Paggio & Navarretta, 2013), and Japanese (Ishi et al., 2014). The most frequently used unimodal gestural micro-feedback signalling sufficient understanding includes nods, nod, smile, up-nod, up-nods, head shakes, head tilt, and eyebrow rise. The top five multimodal micro-feedback expressions signalling sufficient understanding are *yeah+nods, chuckle, yeah+nod, m+nods, and laughter*. Many of them are not only used to communicate CPU but also express the emotions and attitudes of certainty, confirmation, sympathy, surprise, amusement, and interest as well as the evaluative opinion of agreement. Besides these, unimodal head movements are found to exclusively signal sufficient understanding.

In relation to misunderstanding, the data studied show that unimodal vocal-verbal micro-feedback expressions have occurred three times; they are *eh yeah eh, yeah*, and a participant's name, which are usually expressed with hesitation or uncertainty. No unimodal gestural micro-feedback expression occurs. The participants tend to use a multimodal micro-feedback expression that consists of a repetition of the perceived vocal-verbal message (usually a noun phrase) and an assertive gesture nod for information confirmation related to misunderstanding. Half the multimodal expressions contain nods together with *yeah* or a noun phrase. As discussed earlier, assertive micro-feedback expressions such as *yeah* and nod are also sometimes related to misunderstanding. In such cases, the interlocutor always believes he or she has understood the information communicated. However, it usually turns out not to be true sooner or later in further contexts. Besides, a misunderstanding can result in further misunderstandings between the interlocutors back and forth in the dialogue.

Regarding non-understanding, in the data, two unimodal vocal-verbal micro-feedback expressions are used, *sorry* and *what do you mean*, which communicate uncertainty and elicit further information from the other interlocutor. Two unimodal gestural micro-feedback expressions have also occurred, which are eyebrow rise and gaze at. The most frequently used multimodal expressions consist of eyebrow rise or frown, head forward, and gaze sideways together with *sorry*, *what*, or *huh*. The interlocutors use them to express uncertainty, hesitation, and thoughtfulness about the perceived information, and interest in and eagerness to know more. Sometimes, a multimodal micro-feedback chuckle and smile are also used to signal non-understanding, through which the interlocutors often express politeness and embarrassment.

Furthermore, concerning how different modalities of micro-feedback are related to sufficient understanding, misunderstanding, and non-understanding, there were some differences between the Swedish and the Chinese participants that emerged from the data studied. For instance, the Chinese participants seemed to be more expressive in using unimodal vocal-verbal micro-feedback than the Swedes; only they used unimodal vocal-verbal micro-feedback in relation to misunderstanding and non-understanding in the studied data. Equally important, the Swedish participants were found more expressive when using multimodal micro-feedback to signal sufficient understanding in the data. They used unimodal vocal-verbal micro-feedback only to express sufficient understanding, and they had a tendency to use more up-nod(s) as the gestural component for multimodal micro-feedback than the Chinese. These findings in the current study cannot be generalised for Chinese and Swedish speakers in general; rather, they can be only applied to Chinese-Swedish intercultural spontaneous communications to various extents.



# 6

## Results of prosody analysis

In this chapter, based on the empirical interaction data, the prosodic features of the vocal-verbal micro-feedback, that is, pitch, pitch range, pitch contour, and duration, will be investigated in relation to sufficient understanding, misunderstanding, and non-understanding.

The following research question will be addressed. What specific prosodic features of vocal-verbal micro-feedback are correlated to sufficient understanding, misunderstanding, and non-understanding?

### 6.1 Overview of the prosodic data

As presented in Section 4.1.7 earlier, 661 instances of vocal-verbal micro-feedback, which do not overlap but have sufficiently good sound quality, were selected from the entire data (1,288 micro-feedback expressions, see also Chapter 5). Among those 661 vocal-verbal micro-feedback expressions, there are 634, 9, and 18, respectively, related to sufficient understanding, misunderstanding, and non-understanding (see Table 6.1).

Table 6.1. Overview of the prosodic data (Length is presented by means of minute:second, Dial. = dialogue, No. = number, U = understanding).

Dial. No.	Length	No. of Words	No. of Clips	Sufficient U	MisU	Non-U
1	11:44	2,070	101	99	2	0
2	07:56	1,380	80	78	1	1
3	09:04	1,309	99	98	0	1
4	10:29	1,555	70	68	1	1
5	06:52	1,070	79	74	1	4
6	08:11	1,122	95	88	2	5
7	06:08	943	68	63	1	4
8	04:44	678	69	66	1	2
Total	65:08	10,127	661	634	9	18

In order to determine whether there is any association between the prosodic features of the vocal-verbal micro-feedback expressions and the different types of understandings (i.e., sufficient understanding, misunderstanding, and non-understanding) and (if any) what they are, the 661 prosody clips studied were segmented and processed by WaveLab and Praat and annotated manually by independent annotators. Details regarding the research methods, the coding schemes, the coding and annotating procedures, and the inter-coder agreement have been presented in Section 4.1. In the following sections, pitch, pitch range, pitch contour, and duration of the prosodic data will be examined in detail in relation to sufficient understanding, misunderstanding, and non-understanding.

## 6.2 Pitch and duration in relation to understanding

I will start by investigating whether there is any association between the different pitch values or duration values and the types of understandings by means of both distribution box plotting and statistical significance testing.

### 6.2.1 Distribution of pitch and duration values in box plots

Figures 6.1 to 6.4 show *box plots*<sup>48</sup> of the distributions of maximum pitch, minimum pitch, mean pitch, and duration values in relation to different types of understandings.

<sup>48</sup> These box plots are generated automatically by SPSS. A brief explanation of the box plot is presented as follows. The black line (within the box) is the median. The bottom of the box indicates the 25th percentile. Twenty-five percent of clips have values below the 25th percentile. The top of the box represents the 75th percentile. Twenty-five percent of clips have values above the 75th percentile. This means that

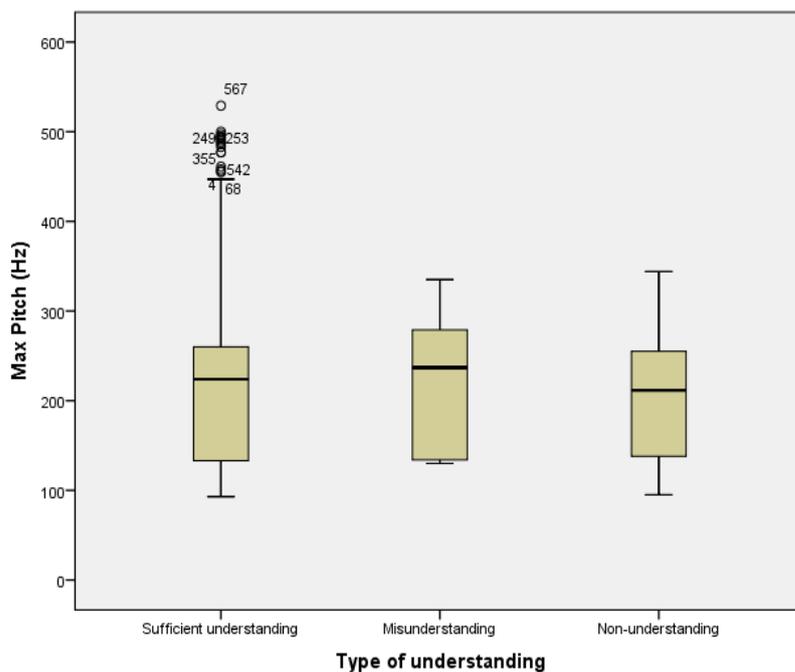


Figure 6.1. Distribution of maximum pitch.

---

50% of the clips lie within the box. The T-bars that extend from the boxes are called inner fences or whiskers. These extend to 1.5 times the height of the box or, if no case/row has a value in that range, to the minimum or maximum values. If the data are distributed normally, approximately 95% of the data are expected to lie between the inner fences. The points are outliers. Numbers next to points show which row it is in the SPSS data list. These are defined as values that do not fall in the inner fences. Outliers are extreme values. The asterisks or stars are extreme outliers. These represent clips that have values more than three times the height of the boxes.

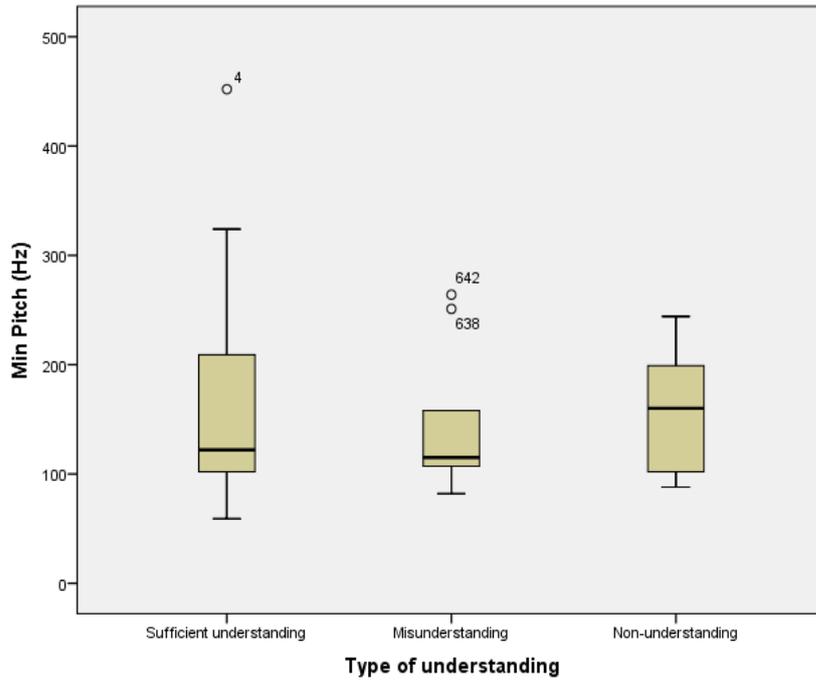


Figure 6.2. Distribution of minimum pitch.

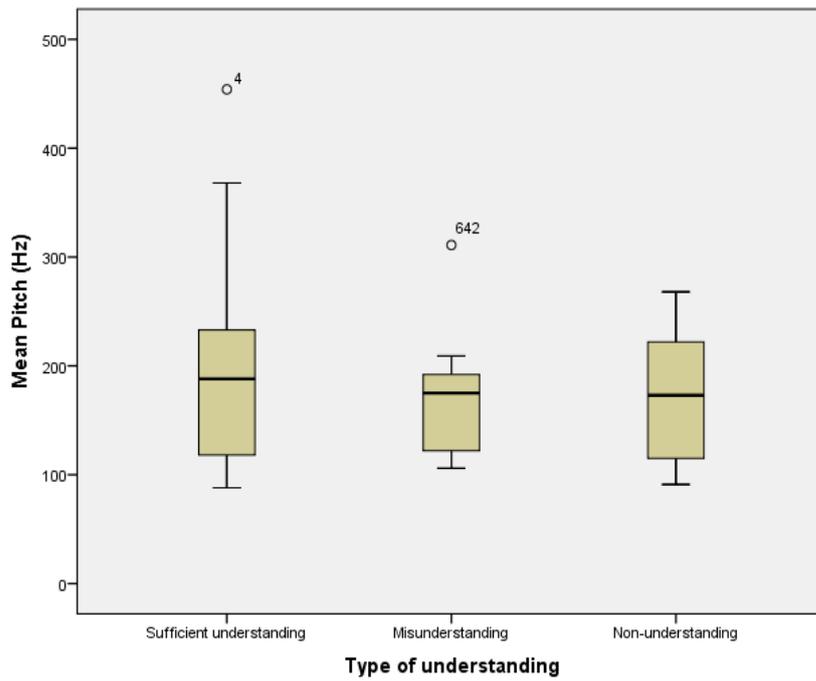


Figure 6.3. Distribution of mean pitch.

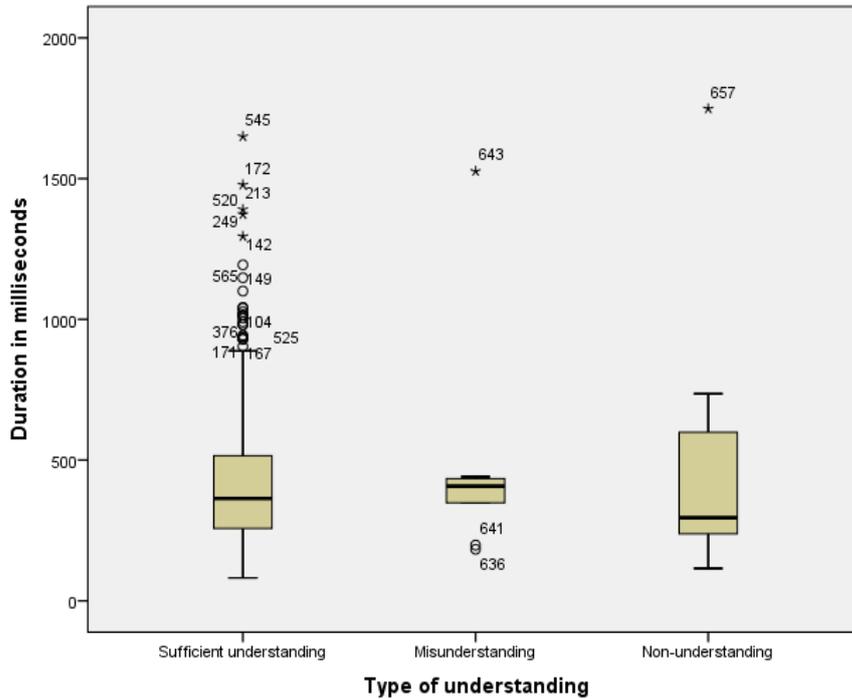


Figure 6.4. Distribution of duration.

As can be seen in each figure above, there are some small differences between the three types of understandings in terms of the plotting of the distributions of pitch and duration values. Concerning the pitch values, for example, as shown from Figures 6.1 to 6.3, the median (i.e., the blacker line within the box) and the bottom line of the box (i.e., the 25th percentile) in the distribution boxes of maximum pitch, minimum pitch, and mean pitch do not differ very much in each figure in relation to sufficient understanding, misunderstanding, and non-understanding. Nevertheless, there are some minor differences in the top line of the box between misunderstanding and the other two types of understandings in the distribution boxes of minimum pitch and mean pitch. Also, with respect to the distribution of duration values, as presented in Figure 6.4, the median does not vary much among the three types of understandings. However, misunderstanding seems to have much smaller variations in duration than sufficient understanding and non-understanding. The top and the bottom lines of the box (i.e., the 75th and the 25th percentile) in misunderstanding show some differences from those in the other two types of understandings.

In order to see if these perceived differences are significant or not, a number of statistical tests are conducted as follows.

## 6.2.2 Kruskal–Wallis statistical tests on pitch and duration values

According to Howell (2010), because the distributions of the studied pitch and duration values are non-normal (e.g., the distribution of pitch values is bimodal and that of duration values is positively skewed), the Kruskal–Wallis one-way analysis of variance is applied in this study. Four pairs of hypotheses are tested:

- (1) H<sub>0</sub>: the distribution of maximum pitch is the same across different types of understandings.  
H<sub>1</sub>: the distribution of maximum pitch is different across different types of understandings.
- (2) H<sub>0</sub>: the distribution of minimum pitch is the same across different types of understandings.  
H<sub>1</sub>: the distribution of minimum pitch is different across different types of understandings.
- (3) H<sub>0</sub>: the distribution of mean pitch is the same across different types of understandings.  
H<sub>1</sub>: the distribution of mean pitch is different across different types of understandings.
- (4) H<sub>0</sub>: the distribution of duration is the same across different types of understandings.  
H<sub>1</sub>: the distribution of duration is different across different types of understandings.

As the Kruskal–Wallis statistical tests reveal (see Tables 6.2 and 6.3), there are no significant differences across the types of understandings in terms of maximum pitch ( $p = .871$ ), minimum pitch ( $p = .952$ ), mean pitch ( $p = .825$ ), and duration ( $p = .569$ ). Because none of them is below the significance level  $\alpha = .05$ , all the four null hypotheses cannot be rejected.

Table 6.2. Summary of the hypothesis tests ( $N$  = number of instances).

	Type of Understanding	$N$	Mean Rank <sup>49</sup>
Max Pitch (Hz)	Sufficient understanding	634	330.76
	Misunderstanding	9	362.78
	Non-understanding	18	323.72
	Total	661	
Min Pitch (Hz)	Sufficient understanding	634	330.58
	Misunderstanding	9	332.83
	Non-understanding	18	344.83
	Total	661	
Mean Pitch (Hz)	Sufficient understanding	634	331.90
	Misunderstanding	9	299.11
	Non-understanding	18	315.42
	Total	661	
Duration (milliseconds)	Sufficient understanding	634	332.06
	Misunderstanding	9	347.78
	Non-understanding	18	285.17
	Total	661	

Table 6.3. The  $p$  values of the Kruskal Wallis Test of statistical significance (Grouping Variable: Type of understanding,  $df$  = degrees of freedom, Asymp. Sig. = asymptotic significance, significance level  $\alpha$  = .05, ms = milliseconds).

	Max Pitch (Hz)	Min Pitch (Hz)	Mean Pitch (Hz)	Duration (ms)
Chi-Square	.276	.098	.385	1.126
$df$	2	2	2	2
Asymp. Sig.	.871	.952	.825	.569

<sup>49</sup> The meaning of “mean rank”: the Kruskal–Wallis test is not concerned with the actual values but their relative ordering. If you run one of these tests, SPSS will rank your scores from lowest (which will be 1) to the highest. You would expect the group with the higher raw values to also have the higher mean rank. For example, if you have 5 males and 5 females with a score on some variable: Male<sub>1</sub> = 2, Male<sub>2</sub> = 3, Male<sub>3</sub> = 5, Male<sub>4</sub> = 8, Male<sub>5</sub> = 11, Female<sub>1</sub> = 13, Female<sub>2</sub> = 15, Female<sub>3</sub> = 20, Female<sub>4</sub> = 21, Female<sub>5</sub> = 22, these scores will be converted to ranks: Male<sub>1</sub> = 1, Male<sub>2</sub> = 2, Male<sub>3</sub> = 3, Male<sub>4</sub> = 4, Male<sub>5</sub> = 5, Female<sub>1</sub> = 6, Female<sub>2</sub> = 7, Female<sub>3</sub> = 8, Female<sub>4</sub> = 9, Female<sub>5</sub> = 10. SPSS will then give you the mean rank and sum of ranks for each group, e.g., for males  $(1+2+3+4+5)/5 = 3$  thus the mean rank for males is 3. Accordingly, the statistics, for instance, in SPSS, will be calculated in relation to the mean rank and the sum of the ranks.

### 6.2.3 ANOVA test on duration

Because the distribution of the duration values (in milliseconds) is non-normal (i.e., positively skewed), the duration data studied are further transformed into normality by taking the logarithm of the values after which the ANOVA test is performed. The purpose is to make one more attempt to test if there is any significant difference between sufficient understanding, misunderstanding, and non-understanding in terms of the duration values of vocal-verbal micro-feedback. As a result, the ANOVA test shows no difference between the three types of understandings, with  $F(2, 658) = 0.327$  and  $p = .721$  (significance level  $\alpha = .05$ ).

To summarise, both Kruskal–Wallis and ANOVA statistical tests show that pitch and duration values are not associated with any specific type of understanding.

## 6.3 Pitch range related to understanding

As presented earlier in Section 4.1.9, the pitch range (measured in Hz) of each micro-feedback is the maximum pitch value minus the minimum pitch value. The maximum and minimum pitch values vary from one micro-feedback item to another, thus providing variable data of pitch ranges. In this section, these pitch range values and pitch range types are studied in relation to the three types of understandings.

### 6.3.1 Relation between pitch range and understanding

In order to investigate the association between pitch range and understanding, the distribution of pitch range values is plotted and compared across the three types of understandings. As shown in Figure 6.5, most of the three types of understandings are expressed with small pitch ranges (mostly below 100 Hz, as can be seen from Figure 6.5). Especially non-understanding has an even smaller pitch range than sufficient understanding and misunderstanding.

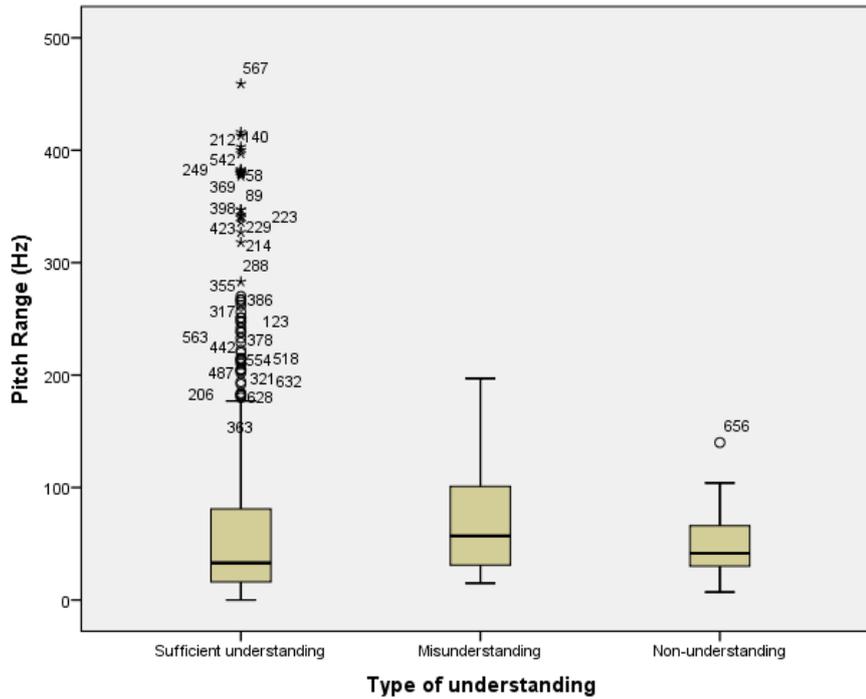


Figure 6.5. Boxplot of distribution of pitch range values for the three types of understandings.

In order to test whether there is any statistical association between pitch range values and the three different understandings (i.e., sufficient understanding, misunderstanding, and non-understanding), the Kruskal–Wallis one-way analysis of variance is used again because of the positively skewed distribution of the data (Howell, 2010). The hypotheses (see below) are examined.

$H_0$ : the distribution of pitch range value is the same across different types of understandings.

$H_1$ : the distribution of pitch range value is different across different types of understandings.

Table 6.4. Summary of the hypothesis test ( $N$  = number of cases).

	Type of Understanding	$N$	Mean Rank
Pitch Range (Hz)	Sufficient understanding	634	328.92
	Misunderstanding	9	409.33
	Non-understanding	18	365.00
	Total	661	

Table 6.5. The  $p$  values of the Kruskal Wallis statistical significance test (Grouping Variable: Type of understanding,  $df$  = degrees of freedom, Asymp. Sig. = asymptotic significance, significance level  $\alpha = .05$ ).

	Pitch Range (Hz)
Chi-Square	2.160
$df$	2
Asymp. Sig.	.340

As presented in Tables 6.4 and 6.5, results show that the null hypothesis cannot be rejected. There is no significant difference ( $H(2) = 2.16, p = .340$ ) in pitch range values across the three types of understandings.

### 6.3.2 Pitch range type in relation to understanding

Since pitch range values have no association with understanding, they are categorised into three pitch range types, that is, small (0–153 Hz), medium (154–306 Hz), and large (307–459 Hz). Details of the categorisation standard and the coding procedure have been presented earlier in Section 4.1.9. In this section, whether there is any specific association between pitch range type and understanding type will be investigated.

Table 6.6. Pitch range type in relation to understanding type (Type of Understanding \* Pitch Range Type cross tabulation).

		Pitch Range Type			
		Small	Medium	Large	Total
Type of Understanding	Sufficient understanding	551	61	22	634
	Misunderstanding	8	1	0	9
	Non-understanding	18	0	0	18
Total		577	62	22	661

As shown in Table 6.6, sufficient understanding, misunderstanding, and non-understanding are all mostly expressed by vocal-verbal micro-feedback with a small pitch range. Fisher's Exact Test 3x3 (two-tailed<sup>50</sup>) results in  $p = .645$ , which is above the

<sup>50</sup> A brief description of the two-tailed and one-tailed statistical tests: in statistical significance testing, the one-tailed test and the two-tailed test are alternative ways of computing the statistical significance of

significance level ( $\alpha = .05$ ). Thus, no further association is found between pitch range type and understanding.

To summarise, pitch range values are not associated with any type of understanding, but all three types of understandings are most frequently related to micro-feedback of the small pitch range type.

## 6.4 Association between duration type and understanding

As tested in the previous section, duration values have no association with any type of understanding. Therefore, the duration values are categorised into three types, that is, short (82–637 milliseconds), medium (638–1192 milliseconds), and long (1193–1748 milliseconds). Details of how the duration data are categorised and coded were presented in Section 4.1.9. In this section, whether any specific type of understanding is associated with any specific duration type and (if any) how they are associated will be investigated.

### 6.4.1 Fisher’s Exact Test (3x3) to examine whether there is any association

Table 6.7 shows that most of sufficient understanding, misunderstanding, and non-understanding have short durations. Fewer of them have medium durations, and very few have long ones.

*Table 6.7.* Overview of duration types in relation to understandings (Type of Understanding \* Duration Type cross tabulation).

		Duration Type			Total
		Short	Medium	Long	
Type of Understanding	Sufficient understanding	541	87	6	634
	Misunderstanding	8	0	1	9
	Non-understanding	14	3	1	18
Total		563	90	8	661

a data set in terms of a test statistic, depending on whether only one direction is considered extreme (and unlikely) or both directions are considered equally likely. Alternative names are one-sided and two-sided tests. The terminology “tail” is used because the extremes of distribution are often small.

The results of Fisher’s Exact Test 3x3 (two-tailed) ( $p = .049$ ,  $\alpha = .05$ ) show that the duration types and the three different understandings (i.e., sufficient understanding, misunderstanding, and non-understanding) are associated. Accordingly, how they are associated will be examined in the next section.

#### 6.4.2 Statistical Fisher’s Exact Test (1x2) to investigate how it is associated

According to Howell (2010), in order to see how duration type and understanding are associated, pairwise comparisons between the frequencies of the short, medium, and long durations are performed separately for the three types of understandings by using Fisher’s Exact Test on a 1x2 table (one-tailed<sup>51</sup>). For instance, for sufficient understanding, the frequencies of short and medium durations are compared first, and then the frequencies of short versus long and medium durations are compared to long durations (see Table 6.8). In this way, nine comparisons are made in all.

*Table 6.8.* Summary of pairwise tests of significant differences in the three duration types in relation to sufficient understanding, misunderstanding, and non-understanding (significance level  $\alpha = .05$ , ns = no significance).

Type of Understanding	Short Duration	Medium Duration	Long Duration	Significance
Sufficient understanding	541	87		$p < .001$
	541		6	$p < .001$
		87	6	$p < .001$
Misunderstanding	8	0		$p = .004$
	8		1	$p = .02$
		0	1	ns ( $p = .5$ )
Non-understanding	14	3		$p = .006$
	14		1	$p < .001$
		3	1	ns ( $p = .313$ )

<sup>51</sup> It was decided to use a one-tailed test, because the study aims to test whether one specific type of duration is more frequent than the other in association with understanding. According to Howell (2010), one-tailed and two-tailed tests in Fisher’s Exact Test are not clear-cut. As he put it, “Fisher’s Exact Test also leads to controversy because of the issue of one-tailed versus two-tailed tests, and what outcomes would constitute a “more extreme” result in the opposite tail. Instead of going into how to determine what is a more extreme outcome, I will avoid that complication by simply telling you to decide in advance whether you want a one- or a two-tailed test and then report the values given by standard statistical software” (Howell, 2010, p. 148).

As presented in Table 6.8, for sufficient understanding, both short and medium durations have  $p$  values below  $\alpha$  (with  $p < .001$  for each). With respect to misunderstanding, when comparing short and medium durations the  $p$  value is .004 and when comparing short and long duration the  $p$  value is .02, both of which are also below  $\alpha$ . In the same way, non-understanding is found to have  $p$  values below  $\alpha$  in comparisons between both short and medium durations ( $p = .006$ ) and between short and long durations ( $p < .001$ ).

To summarise, both misunderstanding and non-understanding are associated more with short duration than medium or long duration. Sufficient understanding is more associated with both short and medium durations, and thus negatively associated with long durations.

## 6.5 Pitch contour and understanding

As presented earlier in Section 4.1.9, the pitch contour of each vocal-verbal micro-feedback is annotated by independent annotators with a coding scheme of falling, flat, and rising. As calculated by Cohen's Kappa, the inter-coder agreement on pitch contour is 0.72, which is considered good (Fleiss, 1971), substantial (Landis & Koch, 1977), and adequate (Brennan & Prediger, 1981) for research studies (see details in Section 4.1.10). In this section, it will be investigated whether pitch contour and understanding are associated.

### 6.5.1 Is there any association between pitch contour and understanding?

As presented in Table 6.9, more than one third of the sufficient understanding cases are expressed with a flat pitch contour and another third with a falling pitch contour. Most of the misunderstanding instances are related to a falling pitch contour, and most of the non-understanding ones are related to a rising pitch contour.

Based on the data shown in Table 6.9, whether there is any association between the type of pitch contour and the type of understanding is tested by using Fisher's Exact Test  $3 \times 3$  (two-tailed). As a result,  $p < .001$  (significance level  $\alpha = .05$ ) is obtained, which suggests that they are associated in some way.

Table 6.9. The distribution of the type of pitch contour and the type of understanding in the prosody data studied (Type of Understanding \* Type of Pitch Contour cross tabulation).

		Type of Pitch Contour			
		Rising	Flat	Falling	Total
Type of Understanding	Sufficient understanding	89	265	280	634
	Misunderstanding	1	1	7	9
	Non-understanding	14	3	1	18
Total		104	269	288	661

### 6.5.2 How are pitch contour and understanding associated?

In order to test the statistical significance of the association, pairwise comparisons between the frequencies of the three types of pitch contours are conducted separately for each of the three types of understandings by using Fisher's Exact Test 1x2 (one-tailed) (Howell, 2010). For each type of understanding, the frequencies of rising and flat pitch contours are compared first, then those of rising versus falling, and next flat compared to falling. In this way, nine comparisons are made in all. A summary of these significance tests is presented below.

Table 6.10. The pairwise statistical significance tests between pitch contour and understanding (significance level  $\alpha = .05$ , ns = not significant).

Type of Understanding	Rising Contour	Flat Contour	Falling Contour	Significance
Sufficient understanding	89	265		$p < .001$
	89		280	$p < .001$
		265	280	ns ( $p = .274$ )
Misunderstanding	1	1		$p = .75$
	1		7	$p = .035$
		1	7	$p = .035$
Non-understanding	14	3		$p = .006$
	14		1	$p < .001$
		3	1	ns ( $p = .313$ )

As can be seen from Table 6.10, for sufficient understanding, a rising pitch contour is less common than flat ( $p < .001$ ) and falling ( $p < .001$ ) pitch contours. For misunderstanding, a falling pitch contour is more common than both rising and flat (both with  $p = .035$ ). For non-understanding, a rising pitch contour is more common than both flat ( $p = .006$ ) and falling ( $p < .001$ ). Other differences in frequencies are non-significant.

To summarise, sufficient understanding is negatively associated with the rising pitch contour but equally associated with flat and falling pitch contours; misunderstanding is associated with the falling pitch contour; also, non-understanding is associated with the rising pitch contour.

## **6.6 Associations among pitch contour, duration type, and pitch range type**

In this section, whether there is any association among pitch contour, duration type, and pitch range type will be investigated.

### **6.6.1 Association between pitch contour and duration type**

In Sections 6.4 and 6.5, certain associations have been found between specific pitch contour or duration type and specific type of understandings. In this section, whether there is any association between pitch contour and duration type will be examined, and then how they are associated will be tested.

**Are pitch contour and duration type associated?**

As presented in Table 6.11, around half of the vocal-verbal micro-feedback expressions, which have short durations, are expressed with a flat or a falling pitch contour. Most of the medium durations are in relation to a falling pitch contour. Vocal-verbal micro-feedback rarely occurs in long durations, and they are not in particular more frequently associated with any specific type of pitch contour than others (see Table 6.11).

In order to test whether there is any association between the type of pitch contour and the type of duration, Fisher's Exact Test 3x3 (two-tailed) is performed. As a result,  $p = .012$  is obtained, which is below the significance level  $\alpha = .05$ . This suggests that pitch contour is associated with duration type of the vocal-verbal micro-feedback.

Table 6.11. Overview of the data in terms of the type of pitch contour and the type of duration (Duration Type \* Type of Pitch Contour cross tabulation).

		Type of Pitch Contour			Total
		Rising	Flat	Falling	
Duration Type	Short	85	242	236	563
	Medium	17	23	50	90
	Long	2	4	2	8
Total		104	269	288	661

In what way are they associated?

In order to test for statistical significance of the association between pitch contour and duration type, pairwise comparisons between the frequencies of the rising, flat, and falling pitch contours are performed separately for the three duration types by using Fisher's Exact Test 1x2 (one-tailed) (according to Howell, 2010). Again, nine comparisons in all are made (see Table 6.12).

Table 6.12. Summary of the pairwise tests of significant differences between the types of durations in relation to the three types of understandings (significance level  $\alpha = .05$ , ns = not significant).

Duration Type	Rising Contour	Flat Contour	Falling Contour	Significance
Short	85	242		$p < .001$
	85		236	$p < .001$
		242	236	ns ( $p = .41$ )
Medium	17	23		ns ( $p = .215$ )
	17		50	$p < .001$
		23	50	$p = .001$
Long	2	4		ns ( $p = .344$ )
	2		2	ns ( $p = .688$ )
		4	2	ns ( $p = .344$ )

Table 6.12 reveals four significant differences with the  $p$  value being either  $< .001$  or  $= .001$ , with significant level  $\alpha = .05$ . First, vocal-verbal micro-feedback of short durations are expressed more with either a flat ( $p < .001$ ) or a falling ( $p < .001$ ) contour than a rising one. Then, the medium duration is associated more with a falling pitch contour (with  $p < .001$  and  $p = .001$ ) than a rising or a flat one. Next, long duration

is found to have no association with any pitch contour type since none of the  $p$  values indicate significance (see “ns” results in Table 6.12).

To summarise, short duration is associated with flat and falling pitch contours, and is thus negatively associated with the rising pitch contour. Medium duration is associated with the falling pitch contour; whereas, no association is present for a long duration.

## 6.6.2 Association between duration type and pitch range type

In order to examine if there is any association between the duration type (i.e., short, medium, and long) and the pitch range type (i.e., small, medium, and large) of the vocal-verbal micro-feedback, first a cross tabulation will be presented and then a statistical account will be provided.

Is there any association between duration type and pitch range type?

As can be seen from Table 6.13, most of the vocal-verbal micro-feedback expressions with a small pitch range have a short duration, and few of them with a large pitch range have a long duration.

*Table 6.13.* Distribution of the data in terms of pitch range type and duration type (Duration Type \* Pitch Range Type cross tabulation).

		Pitch Range Type			Total
		Small	Medium	Large	
Duration Type	Short	504	45	14	563
	Medium	68	15	7	90
	Long	5	2	1	8
Total		577	62	22	661

With the help of Fisher’s Exact Test 3x3 (two-tailed), the result ( $p = .001$ ,  $\alpha = .05$ ) shows that duration type is associated with pitch range type. In order to see how they are associated, statistical tests are carried out as follows.

How is the association?

In order to test for statistical significance of the association between duration type and pitch range type, pairwise comparisons between the frequencies of the pitch

range types are performed separately for the three duration types by using Fisher's Exact Test 1x2 (one-tailed).

*Table 6.14.* Summary of pairwise tests of significant differences in frequencies for pitch range type in relation to each duration type (significance level  $\alpha = .05$ , ns = not significant).

Duration Type	Pitch Range Type			Significance
	Small	Medium	Large	
Short	504	45		$p < .001$
	504		14	$p < .001$
		45	14	$p < .001$
Medium	68	15		$p < .001$
	68		7	$p < .001$
		15	7	ns ( $p = .067$ )
Long	5	2		ns ( $p = .227$ )
	5		1	ns ( $p = .109$ )
		2	1	ns ( $p = .75$ )

The results in Table 6.14 show that both the short and the medium durations are associated with a small pitch range for micro-feedback. Due to low frequencies, no association is found for long duration.

### 6.6.3 Association between pitch range type and pitch contour

The relation between the pitch range type and the pitch contour of the vocal-verbal micro-feedback will be investigated in this section.

Is pitch range type associated with pitch contour?

An overview of how the prosodic data studied is distributed in terms of the pitch range type and the pitch contour is presented in Table 6.15. As shown, most of the micro-feedback items expressed with a small pitch range have either a flat or a falling pitch contour. Most of those with a medium pitch range have a falling pitch contour. Few of them with a large pitch range have a rising pitch contour.

Table 6.15. Distribution of the data in pitch range type and pitch contour (Pitch Range Type \* Type of Pitch Contour cross tabulation).

		Type of Pitch Contour			Total
		Rising	Flat	Falling	
Pitch Range Type	Small	87	249	241	577
	Medium	15	15	32	62
	Large	2	5	15	22
Total		104	269	288	661

Is any specific pitch range type associated more than others with any specific type of pitch contour? Fisher's Exact Test 3x3 (two-tailed) result ( $p = .005$ ,  $\alpha = .05$ ) shows that there is some association.

How are they associated?

In order to test for the statistical significance of the association between pitch range type and pitch contour, pairwise comparisons between the frequencies of the pitch contour (three types) are performed separately for the three pitch range types, by using Fisher's Exact Test 1x2 (one-tailed).

Table 6.16. Summary of pairwise tests of significant differences in frequencies for type of pitch contour for each pitch range type (significance level  $\alpha = .05$ , ns = not significant).

Pitch Range Type	Type of Pitch Contour			Significance
	Rising	Flat	Falling	
Small	87	249		$p < .001$
	87		241	$p < .001$
		249	241	ns ( $p = .376$ )
Medium	15	15		ns ( $p = .572$ )
	15		32	$p = .009$
		15	32	$p = .009$
Large	2	5		ns ( $p = .227$ )
	2		15	$p = .001$
		5	15	$p = .021$

Table 6.16 shows that the  $p$  values are significant when comparing the rising pitch contour to the flat pitch contour ( $p < .001$ ) and also when comparing it to the falling

pitch contour ( $p < .001$ ) within the same small pitch range. Similarly, when it comes to the medium pitch range, there are significant differences between the rising and the falling pitch contours ( $p = .009$ ) and between the flat and the falling ones ( $p = .009$ ). With respect to the large pitch range, the difference between the rising and the falling pitch contours ( $p = .001$ ) as well as that between the flat and the falling pitch contours ( $p = .021$ ) seem to be significant.

To summarise, for small pitch range, flat and falling pitch contours both occur to the same degree more frequently than the rising pitch contour. Thus, small pitch range is negatively associated with a rising pitch contour. Medium and large pitch ranges are more associated with the falling pitch contour than others.

#### 6.6.4 Summary of the inter-associations

The inter-associations among pitch contour, duration type, and pitch range type can be presented as follows (see Table 6.17).

Table 6.17. Inter-associations among pitch contour, duration type, and pitch range type.

Pitch Contour	Duration Type	Pitch Range Type
falling contour	short and medium durations	small, medium, and large pitch ranges
flat contour	short duration	small pitch range
rising contour	negatively with short duration	negatively with small pitch range

Short duration is associated with flat and falling pitch contours of micro-feedback thus negatively associated with a rising pitch contour. Medium duration is associated with a falling pitch contour whereas no association is present for long duration. Both short and medium durations are associated with a small pitch range. No association is found for long duration. Small pitch range is associated to the same degree with the flat and falling pitch contours, and it is negatively associated with a rising pitch contour. Medium and large pitch ranges are both associated with the falling pitch contour.

To summarise, the falling pitch contour accompanies both short and medium durations in all types of pitch ranges. The flat pitch contour is associated with the short duration and the small pitch range. The rising pitch contour is negatively associated with the short duration and the small pitch range.

## 6.7 Discussion

The present analysis primarily deals with the prosodic properties of the perceived and understood micro-feedback in speech forms, and examines the acoustic correlates of pitch and duration. This approach to studying prosody of spoken language is in line with many researchers such as Pollack et al. (1960), Crystal (1969), Nöth et al. (2002), and Patel and Grigos (2006).

### 6.7.1 On pitch values, pitch range values, and pitch range types

It has been found that pitch values and pitch range values are not associated with any specific type of understanding. This is probably because the data studied involves eight participants and the individual differences are prominent when studying pitch features in the conversation (Couper-Kuhlen, 1996). People have various basic pitch levels. Moreover, these eight participants come from two dramatically different cultures, Swedish and Chinese, with two males and two females for each cultural group. Individuals in even the same culture have different phonemic characteristics. The cultural, gender, and individual differences very likely exacerbate the difficulties in processing the pitch cues.

In the present analysis, there are equal numbers of Swedish and Chinese as well as male and female participants, and the prosodic features all participants have in common are of research interest. Without normalising the prosody data, the current study decreases the chance of finding differences, however, increases the variety of the data and probably also increases the validity of the analysis (see more discussions in Section 4.1.7).

All the three types of understandings, that is, sufficient understanding, misunderstanding, and non-understanding, have been found to be most commonly expressed by vocal-verbal micro-feedback with a small pitch range (0 to 153 Hz). This is probably because most of the vocal-verbal micro-feedback expressions are short words like *yeah*, *okay*, *yes*, *no*, *m*, *oh*, *ah*, *mhm*, *uhu* (see also Jurafsky et al., 1998; Ward & Tsukahara, 2000; Lu & Allwood, 2011) with reasonably short durations (see more discussion in the following section). According to the positive association between pitch range and duration (Xu & Wang, 2009), these most commonly used micro-feedback words are of short duration and have a relatively small pitch range (e.g., nearly below 150 Hz).

In addition, most of the vocal-verbal micro-feedback items studied express CPUE/A (i.e., contact, perception, understanding, and emotion/attitude) through their prosody and other accompanying information such as gesture. The complementary role of prosody in the vocal-verbal information, for instance, by adding

emotions and attitudes has been found in a considerable number of the vocal-verbal micro-feedback expressions, for example, of both small and large pitch ranges. As presented in Chapter 2, these communicative functions of micro-feedback are interpreted and annotated depending on the context (e.g., contextualisation in particular with a focus on context dependency) (see discussions in Gumperz, 1982; Bauman & Briggs, 1990; Tannen, 1993; Couper-Kuhlen, 2001) and the way micro-feedback is communicated (e.g., modality and prosody features). As presented in the analysis, vocal-verbal micro-feedback with both small and large pitch ranges play important roles in communication. In particular, vocal-verbal micro-feedback expressions with small pitch ranges are associated with all types of understandings. Thus, Zuraidah and Knowles' (2006) idea that the prosody of a narrow pitch range (in the Malay language) does not play any significant role in communication because the small pitch range makes the prosodic patterns difficult to identify is not confirmed in the studied Swedish–Chinese data (in the English language).

These findings suggest that pitch values, pitch range values, and pitch range types of speech can possibly reflect some more interesting patterns than what has been already discussed, when culture, gender, and individual differences as well as variations of the word or vowel length are taken into account. In other words, in order to discover whether there is any association between pitch and understanding in communication and how it is associated, preferred data may be the prosodic features of a particular word or vowel of one specific individual speaker (taking into consideration culture and gender factors).

### 6.7.2 Duration type and understanding

As discussed (in Section 4.1.7), individuals have various rates of speech in general. Some speak faster and others speak more slowly. Even a single individual can vary speaking rates in different situations (e.g., with an interesting vs. less interesting topic, or in relaxed vs. stressed mood). It is not easy to generalise the prosodic feature of duration. There are, nevertheless, some indicative results presented in the data.

Sufficient understanding has been found to be associated with micro-feedback of both short and medium durations, thus negatively associated with long duration. Both misunderstanding and non-understanding are associated with micro-feedback with a short duration. This is probably because most of the vocal-verbal micro-feedback expressions are single words rather than long phrases or sentences (as discussed earlier).

The findings suggest that vocal-verbal micro-feedback of short and medium durations can reveal more information about understanding than those of long dura-

tions. In other words, from short and medium vocal-verbal micro-feedback expressions, people are very likely to acquire knowledge if the information has been sufficiently understood, misunderstood, or not understood.

Intuitively, it may be assumed that messages that have a longer duration may contain more information. However, according to the results that none of the three types of understandings is positively associated with long duration, this intuitive assumption is not confirmed. Instead, the results show that speech forms (such as vocal-verbal micro-feedback) of short and medium durations can still contain rich information in relation to understanding and sense-making.

### 6.7.3 Pitch contour type in relation to understanding

It is not uncommon that people tend to converge or adapt their communicative behaviours when they communicate with one another (Allwood & Lu, 2011), including pitch levels and tones. Also, Chinese and Swedish are two languages that use prosody phonemically, and the participants were speaking their common language, the English lingua franca, in which prosody does not play the same role and is normally influenced by the speaker's first language. The findings regarding contour features presented in the analysis can be very specific to this particular data.

It has been found that sufficient understanding is negatively associated with a rising pitch contour but positively associated with the flat and the falling pitch contours. Another finding is that misunderstanding is related more to a falling pitch contour than others. These results turn out to be in line with a number of studies such as those by Patel and Grigos (2006) and Zuraidah and Knowles (2006), in which a falling or a flat pitch contour is more frequently used in statements than a rising pitch contour. Obviously, both sufficient understanding and misunderstanding are more commonly expressed or embedded in statements than other speech acts. It is common that when the interlocutors' responses show sufficient understanding and misunderstanding, the interlocutors are stating what they have understood and believed. This is also compatible with what House (2006) said, that misunderstanding often occurs when the hearers perceive the cues consistent with those they are familiar with, for example, a falling pitch contour. Thus, how to distinguish misunderstanding from sufficient understanding by merely the micro-feedback linguistic form itself or its prosodic cues such as the "daily life familiar" (based on House's "familiar" above) falling pitch contours is not obvious. As a matter of fact, more factors like the interaction context and the interaction (interrelation) between vocal-verbal (including prosody) and gestural components of micro-feedback or other linguistic phenomena all play roles.

On the other hand, for requesting, for example, asking a question or asking for further clarification, Patel and Grigos (2006), Zuraidah and Knowles (2006), and

House (2006) have suggested that a rising pitch contour is more frequently in use. This is confirmed in the present analysis, because the results show that non-understanding is usually communicated in a rising pitch contour and usually through questions consisting of vocal-verbal micro-feedback like *sorry*, *what*, *what do you mean*, and *what did you say* (see Chapter 5 for more details). Also, these are consistent with Kushida's (2011) statement that when people repair the conversation (e.g., by rephrasing or paraphrasing), for instance, by means of some micro-feedback items, they usually communicate with an upward (or rising) intonation. The rising (or *acceleration*, as suggested by Schegloff, 1998) prosodic pattern may not serve to project the possibility of completing the utterance. As discussed earlier, such vocal-verbal micro-feedback expressions function as eliciting devices for further clarification of information in the communication construction process. Thus, the association between non-understanding and the micro-feedback of a rising pitch contour is confirmed.

#### 6.7.4 Inter-associations among pitch contour type, duration type, and pitch range type

Results of the empirical study show that pitch contour, duration type, and pitch range type of vocal-verbal micro-feedback are inter-associated with each other.

Regarding pitch contour and duration type (see Table 6.17), short durations have been found to be negatively associated with a rising pitch contour, but positively associated with a flat or a falling pitch contour. Medium durations are associated with a falling pitch contour. No association is present for a long duration. Together with the earlier findings that most of the vocal-verbal micro-feedback expressions are used to signal sufficient understanding and that sufficient understanding is mostly expressed with a falling or a flat pitch contour, a claim can be made that most of the vocal-verbal micro-feedback expressions are used to signal sufficient understanding with a falling or a flat pitch contour, for which the falling pitch contour is used with both short and medium durations and the flat pitch contour only with a short duration. Besides having confirmed Tomlinson Jr and Fox Tree's (2011) finding that long duration is unrelated to the rising pitch, these findings also show that long duration is not associated with any type of pitch contour.

As regards duration type and pitch range type, short and medium durations have been found to be associated with a small pitch range (see also Table 6.17). This is consistent with Xu and Wang's (2009) theory of positive association between pitch range and duration. Also, no association has been found between long duration and any pitch range type. This is possibly because there are far fewer vocal-verbal micro-feedback expressions with long durations than with short and medium durations.

With respect to pitch range type and pitch contour, the small pitch range is negatively associated with the rising pitch contour, but positively associated with the falling and flat pitch contours. Medium and large pitch ranges are associated with the falling pitch contour. In fact, these are logically in line with the findings concerning the associations (as discussed above) between pitch contour and duration type and between duration type and pitch range type.

## **6.8 Conclusion of Chapter 6**

In this chapter, the prosodic features of 661 vocal-verbal micro-feedback expressions in relation to sufficient understanding, misunderstanding, and non-understanding have been investigated. The research question, what specific prosodic features of vocal-verbal micro-feedback are correlated to sufficient understanding, misunderstanding, and non-understanding, has been examined using an explorative approach.

The results show that pitch, pitch range, and duration values are not correlated with any specific type of understanding. This is probably because of individual differences and variations in the word or vowel length. When it comes to data that also involve different variations of culture, gender, and language (first language and current communication language), it is even more difficult to distinguish any general pattern of pitch, pitch range, and duration values. However, the findings that are generalisable may suggest some possible universal patterns for normal social interactions.

As regards the pitch range type and understanding, results show that all types of understandings are most commonly related to the vocal-verbal micro-feedback that has a small pitch range. This finding does not confirm Zuraidah and Knowles' (2006) idea that the prosody of a narrow pitch range makes the prosodic patterns difficult to identify. Accordingly, the small pitch range does not play any significant role in communication.

Regarding duration type and understanding, the statistical tests show that sufficient understanding is positively associated with both short and medium durations, and thus negatively associated with long durations. Both misunderstanding and non-understanding are associated with short duration. This suggests that vocal-verbal micro-feedback with short and medium durations can reveal more information in relation to understanding in general than those with a long duration. This finding does not confirm the intuitive assumption that information of longer duration may contain more information regarding understanding. On the contrary, speech forms, such as vocal-verbal micro-feedback of short and medium duration can play significant roles in understanding and sense-making in social interaction.

With respect to pitch contour and understanding, the results show that sufficient understanding is negatively associated with the rising pitch contour but associated to the same degree with the flat and falling pitch contours. Misunderstanding is related more to a falling pitch contour than other contours, and non-understanding is associated with a rising pitch contour. These findings turned out to be consistent with a number of studies such as those by Patel and Grigos (2006) and Zuraidah and Knowles (2006) in that a falling or a flat pitch contour is more frequently used than a rising pitch contour in statements and most sufficient understanding and misunderstanding cases are expressed or embedded in statements rather than in other speech acts. On the other hand, a rising pitch contour is more frequently in use than other contours for requesting such as asking a question or asking for further clarification, typically as non-understanding occurs. These findings suggest some possible universal patterns that could be further examined in the future.

Besides these, the statistical tests also show that there are specific inter-associations among pitch contour, duration type, and pitch range type (see Table 6.17). The falling pitch contour is correlated with both short and medium durations in all types of pitch ranges. The flat pitch contour is correlated with short duration and small pitch range. The rising pitch contour is negatively associated with short duration and small pitch range.

This empirical study shows that the prosodic features, especially pitch contour and duration type of the uttered micro-feedback, provide clues for recognising sufficient understanding, misunderstanding, and non-understanding. For instance, sufficient understanding is normally expressed in a flat or a falling pitch contour and with a short or medium duration, misunderstanding is usually associated with a falling pitch contour and a short duration, and non-understanding is normally expressed with a rising pitch contour and a short duration. A falling pitch contour accompanies all pitch range types, whereas a flat pitch contour accompanies only a small pitch range.

The current analysis has adopted the widely accepted view that prosody makes a significant and systematic contribution to conversation (see Couper-Kuhlen & Selting, 1996; Ward & Tsukahara, 2000; House, 2006). Also, this study has argued in favour of a contextual account for examining the relations between prosody (of micro-feedback in this particular study) and understanding. Statistical tests suggest that pitch contour, duration type, and pitch range type are all respectively correlated to understanding, and they are also inter-associated with one another. However, the prosodic features of micro-feedback do not give all the sufficient information about understanding in conversation; rather, they provide something necessary to assist in identifying different types of understandings. During this identifying process, contextualised accounts with a particular focus on context dependency play important roles.

# 7

## Discussion and summary of Study 1

In this chapter, first, a review of the research purpose and research questions in Study 1 will be provided. Second, the main empirical findings of the analyses will be summarised. Third, contributions and implications of Study 1 in the thesis will be discussed. Next, I will present critical reflections on the analyses and point out the research limitations of Study 1. This will be followed by suggestions for Study 2.

### 7.1 Review of the research purpose and research questions in Study 1

The general theoretical literature on the subject of micro-feedback in communication was incomplete as regards several vital questions, specifically in relation to understanding in the discourse. Study 1 in the thesis sought to answer three of these questions:

- RQ1: How are the auditory and visual modalities involved in micro-feedback expressions that are related to sufficient understanding, misunderstanding, and non-understanding?
- RQ2: What are the typical unimodal and multimodal micro-feedback expressions that signal sufficient understanding, misunderstanding, and non-understanding?

RQ3: What specific prosodic features of vocal-verbal micro-feedback are correlated to sufficient understanding, misunderstanding, and non-understanding?

RQ1 and RQ2 were investigated in Chapter 5, and RQ3 was addressed in Chapter 6. Study 1 in the thesis was based on eight audio- and video-recorded FTF dyadic intercultural communication dialogues in the English lingua franca between Swedish and Chinese participants who were strangers to each other. Their communication task was to become acquainted with one another.

Study 1 aims to investigate micro-feedback in relation to understanding issues in a spontaneous communication activity in first encounters. Two empirical analyses were conducted to explore the phenomenon of micro-feedback and identify the features of micro-feedback in terms of its modality (in Chapter 5) and prosody (in Chapter 6) in relation to understanding in real-time communication.

## **7.2 Summary of the empirical findings in Study 1**

The empirical findings from the respective empirical chapters are summarised here: Chapter 5 Results of modality analysis and Chapter 6 Results of prosody analysis.

Most of the micro-feedback expressions have been found to signal sufficient understanding, a few to convey non-understanding, and even fewer are related to misunderstanding. This suggests that compared to sufficient understanding and non-understanding, misunderstanding is least related to micro-feedback and is most difficult to observe in spontaneous communication.

This section will synthesise the empirical findings (see Table 7.1) in order to answer the research questions.

*Table 7.1.* A general overview of the empirical results regarding the most frequent micro-feedback expressions and their modality and prosody features (C = contact, P = perception, misP = misperception, -P = non-perception, U = understanding, misU = misunderstanding, -U = non-understanding, E/A = emotional and attitudinal reaction, usu. = usually, vocal micro-feedback words are italicised. *Note.* The expressions and functions presented are sequenced in decreasing frequencies.)

		Sufficient understanding	Misunderstanding	Non-understanding
<b>Modality</b>				
Unimodal	Vocal-verbal	<i>yeah, okay, m, ah, yes, no, yeah yeah yeah, yeah yeah, ah yeah, yeah okay, ah okay, okay okay</i>	<i>eh yeah eh, yeah</i> , a participant's name	sorry, what do you mean
	Gestural	nods, nod, smile, up-nod, up-nods, head shakes, head tilt, eyebrow rise (Note. Head movements exclusively signal it.)	NONE	eyebrow rise, gaze at
Multimodal		<i>yeah+nods, chuckle, yeah+nod, m+nods, laughter, okay+nods, mhm+nod, okay+nod, okay+up-nod, yeah+up-nod, yes+nod, m+nod</i>	half of the misUs contain <i>nod+yeah</i> or <i>nod+a noun phrase</i> (usu. the perceived message) (Note. Because it usu. contains <i>nod</i> and <i>yeah</i> , it is not easy to detect misU. Also, misU can result in further misUs.)	usu. consists of eyebrow rise or frown, head forward, gaze sideways or gaze at plus <i>sorry, what, huh</i> , or simply chuckle, smile
<b>Communicative functions</b>				
	CPU	CPU	CPmisU or CmisPmisU	CP-U or C-P-U
	E/A	certainty, confirmation, sympathy, surprise, amusement, interest	usu. for information confirmation  hesitation, uncertainty	Uncertainty, hesitation, and thoughtfulness about the perceived information, interest and eagerness of knowing more, or simply politeness and embarrassment
	Evaluative opinion	agreement	NONE	NONE
<b>Prosody</b>				
Pitch	Range	small	small	small
	Contour	flat, falling	falling	rising
Duration	Type	short, medium	short	short

### 7.2.1 Regarding modality of micro-feedback in relation to understanding

How are the auditory and visual modalities involved in micro-feedback, and what are the typical unimodal and multimodal micro-feedback, which signals sufficient understanding, misunderstanding, and non-understanding? The empirical analysis of modality presented in Chapter 5 shows that sufficient understanding is signalled through unimodal (vocal-verbal and gestural almost to the same degree) more than multimodal micro-feedback. Misunderstanding involves more multimodal than unimodal micro-feedback, and it is not conveyed through unimodal gesture at all. Non-understanding is mostly expressed by multimodal micro-feedback expressions and rarely through unimodal ones. These findings suggest that in FTF, spontaneous communication unimodal gestural micro-feedback plays an important role in relation to sufficient understanding, but has little to do with non-understanding and nothing to do with misunderstanding. That is, when unimodal gesture is involved, the micro-feedback very likely signals sufficient understanding; otherwise, there is probably some problem in understanding, for example, misunderstanding or non-understanding.

For sufficient understanding, unimodal vocal-verbal micro-feedback *yeah, okay, m, ah, yes, no, yeah yeah yeah, yeah yeah, ah yeah, yeah okay, ah okay, and okay okay*, unimodal gestural micro-feedback nods, nod, smile, up-nod, up-nods, head shakes, head tilt, and eyebrow rise, and multimodal micro-feedback *yeah+nods, chuckle, yeah+nod, m+nods, laughter, okay+nods, mhm+nod, okay+nod, okay+up-nod, yeah+up-nod, yes+nod, and m+nod* are most frequently employed (in a sequence of decreasing frequency). Many of them are not only used to communicate “I hear you, I perceive and understand what you have said, and I would like to continue the conversation with you” (CPU), but also express emotions and attitudes of certainty, confirmation, sympathy, surprise, amusement, and interest as well as the evaluative opinions of agreement and disagreement. Unimodal head movements are found to exclusively signal sufficient understanding.

Regarding misunderstanding, unimodal vocal-verbal micro-feedback *eh yeah eh, yeah*, and a participant’s name are usually expressed with hesitation or uncertainty. No unimodal gestural micro-feedback occurs in relation to misunderstanding. Half the multimodal micro-feedback expressions contain nod together with *yeah* or a noun phrase, which is usually the perceived message and functions as information confirmation. Because these micro-feedback expressions usually involve *yeah* and nod, which are most frequently used to signal sufficient understanding, it is not easy to distinguish misunderstanding from sufficient understanding by merely looking at

its micro-feedback signals. Interactional consideration based on the context of relevance should always be taken into account. For example, *yeah* and nod, which are related to misunderstanding usually have the specific emotions and attitudes such as hesitation and uncertainty communicated through the accompanied prosodic features and/or gestures. Misunderstanding may be found in a richer context. Also, this may result in further misunderstandings between the interlocutors back and forth in the interaction context.

With respect to non-understanding, unimodal vocal-verbal micro-feedback *sorry* and *what do you mean*, unimodal gestural micro-feedback eyebrow rise and gaze at, and multimodal micro-feedback consisting of eyebrow rise or frown, head forward, and gaze sideways or gaze at combined with *sorry*, *what*, or *huh* are frequently used. They often express uncertainty, hesitation, and thoughtfulness about the perceived information and interest and eagerness to know more. With the Swedish and Chinese participants, multimodal micro-feedback chuckle and smile are also found to signal non-understanding, expressing politeness and embarrassment.

### 7.2.2 On prosody of micro-feedback correlated to understanding

What specific prosodic features of vocal-verbal micro-feedback are correlated to sufficient understanding, misunderstanding, and non-understanding? The empirical analysis presented in Chapter 6 focuses on prosody. The results show that there is no correlation between understanding and the pitch, pitch range, and duration values of micro-feedback, probably because of culture, gender, and individual differences as well as variations in the micro-feedback word and vowel length. However, understanding has been found to be correlated to the pitch range type, duration type, and pitch contour of the micro-feedback expression.

As regards pitch range type, sufficient understanding, misunderstanding, and non-understanding are most commonly expressed by vocal-verbal micro-feedback, which has a small pitch range. The large pitch range does not play any significant role in relation to understanding. Regarding duration type, both misunderstanding and non-understanding are associated with short duration. Sufficient understanding is positively associated with both short and medium durations and thus negatively associated with long duration. This suggests that compared to vocal-verbal micro-feedback with a long duration, those with short and medium durations can reveal more information about whether the communicated message has been sufficiently understood, misunderstood, or not understood. Concerning pitch contour, sufficient understanding is negatively associated with a rising pitch contour and associated to the same degree with the flat and falling pitch contours. Misunderstanding is

associated with a falling pitch contour, and non-understanding is usually communicated with a rising pitch contour.

Furthermore, pitch contour, duration type, and pitch range type have been found to be inter-associated with one another. The falling pitch contour is associated with both short and medium durations in all types of pitch ranges. The flat pitch contour is associated with a short duration and a small pitch range. The rising pitch contour is negatively associated with the short duration and the small pitch range.

### 7.2.3 Generalisation

To synthesise the empirical findings of the two analyses, Study 1 in this thesis presents the specific patterns of micro-feedback in terms of modality and prosody in relation to understanding. Sufficient understanding is frequently signalled by unimodal vocal-verbal *yeah*, *okay*, *m*, *ah*, and *yes*, unimodal gestural nods, nod, smile, up-nod, and up-nods, and multimodal *yeah+nods*, chuckle, *yeah+nod*, *m+nods*, laughter, and *okay+nods* (top five for each), of which the vocal-verbal micro-feedback is associated with a small pitch range, a flat or a falling pitch contour, and a short or a medium duration. Misunderstanding normally occurs where *yeah* and nod are usually used with hesitation and uncertainty, with *yeah* associated with a small pitch range, a falling pitch contour, and a short duration. Non-understanding is usually signalled by unimodal vocal-verbal *sorry* and *what do you mean*, unimodal gestural eyebrow rise and gaze at, and multimodal eyebrow rise or frown, head forward, gaze sideways, or gaze at plus *sorry*, *what*, or *huh*, and sometimes chuckle and smile, of which the vocal-verbal micro-feedback is associated with a small pitch range, a rising pitch contour, and a short duration.

Study 1 in this thesis confirms the relation between micro-feedback and understanding, which has been more or less stressed by Clark and Schaefer (1989), McConnell (1993), Ryan and Conover (2004), and Boud and Molloy (2013) (see details in Chapters 1 and 2). Micro-feedback (or the like) is one type of evidence of showing understanding, and it plays an important role in the communication exchange process, which eventually leads to meaning and understanding sharing. Also, empirical results of the second analysis confirm the relation between prosody and understanding, which has been addressed to some extent by Grice (1989), Couper-Kuhlen and Selting (1996), Ward and Tsukahara (2000), House (2006), and Mitchell and Ross (2013) (see also Chapters 1 and 2). Prosody has a pragmatic language function, and the prosodic aspects of speech that supplement or modify the meaning of the spoken word help to understand the speaker's meaning. Equally important, findings from the first empirical analysis suggest that modality has some specific relations to understanding. Here, sufficient understanding is more frequently signalled

through unimodal micro-feedback than multimodal micro-feedback, but misunderstanding and non-understanding are the opposite. Unimodal head movements exclusively signal sufficient understanding. Unimodal gestures are not involved in misunderstanding at all. These results can be strengthened and expanded in future research. Furthermore, Study 1 in this thesis demonstrates the inter-dependency between micro-feedback (both gesture and prosody), understanding, and context, which to varying extents has been suggested with different focuses by Lindwall (2008), Linell (2009), Nadeu and Prieto (2011), Hindmarsh et al. (2011), and Finkbeiner et al. (2012). That is, in order to assess or evaluate understanding in social interaction, various prosodic aspects of speech together with its relevant gestural information as well as the contextual information have to be taken into account. In communication, context, on the one hand, links together the linguistic vocal-verbal, prosodic, and gestural behaviours which, on the other, also help construct the context.

In general, both modality and prosody of micro-feedback play significant and systematic roles in communication and understanding. Both have some specific relations to sufficient understanding, misunderstanding, and non-understanding. They do not provide all the sufficient conditions for identifying and evaluating understanding in conversation, but do provide some necessary conditions and assistance for it. Still, with the contextualised and interactional accounts of relevance, primarily the context dependency consideration, micro-feedback and its modality and prosody features can provide the analyst and the interlocutor with more information about how the communicated message has been understood.

### **7.3 Contributions and implications of Study 1**

Study 1 in this thesis makes theoretical and practical contributions and implications. In this section, how empirical findings and implications in Study 1 are in line with or contradict some earlier theories will be discussed.

#### **7.3.1 The most frequently used unimodal micro-feedback**

Study 1 in this thesis has found that the most common unimodal vocal-verbal micro-feedback expressions are *yeah*, *okay*, *m*, *ah*, *yes*, *no*, *yeah yeah yeah*, *yeah yeah*, *ah yeah*, *yeah okay*, *ah okay*, and *okay okay* (top six for each in a sequence of decreasing frequency). Among them, *yeah*, *okay*, *m*, *ah*, and *yes* are comparable with the five most frequent *backchannel* expressions found in other corpora and studies such as those by Jurafsky et al. (1998) and Ward and Tsukahara (2000). Furthermore, the most frequent unimodal gestural micro-feedback expressions found in the present

study are single and repeated head nods (i.e., nod and nods). This result corresponds well with a few other studies of *feedback* in several languages, that is, Swedish and Finnish (Navarretta et al., 2012), Danish (Paggio & Navarretta, 2013), and Japanese (Ishi et al., 2014).

These unimodal vocal-verbal or gestural micro-feedback expressions are always used to communicate “I hear you, I perceive and understand what you have said, and I would like to continue the conversation with you” (CPU) and sometimes express the emotions and attitudes (E/A) of certainty, confirmation, sympathy, surprise, amusement, and interest as well as the evaluative opinion agreement. These may suggest some possible universal patterns of using micro-feedback in social interaction.

### 7.3.2 Problems and difficulties in understanding in intercultural first encounters

According to Gumperz (1982), Tannen (1990), and Samovar et al. (2012), in a joint communication activity, people who have different cultural and language backgrounds probably have more problems and difficulties understanding than those who have the same relevant backgrounds. Also, the unacquainted people who have mutually distinct and unknown personal and professional experiences, different individual communication presuppositions and expectations, and various limitations in common knowledge and resources in sense-making may have many understanding problems in communication.

The present study, however, shows that out of 1,288 cases of all types of understandings communicated in relation to micro-feedback, the participants have 1,256 cases of sufficient understanding, 9 of misunderstanding, and 23 of non-understanding in all. It seems that there are not as many problems and difficulties in understanding as predicted in this particular Swedish–Chinese intercultural first encounter’s conversational data. This is probably because the participants’ shared social background of being university students in Sweden and their relatively good mastery of the communicative language English, it is not very difficult for them to achieve mutual understanding in general. Of course, it is also possible that people do not want to reveal the problem of understanding very explicitly, so they try to minimise it as much as possible in order to be polite in a socially conventional way and not lose face.

Besides this, it has been found that the Swedes and the Chinese have very similar communication intelligibility. For instance, the Swedish speakers misunderstood the Chinese 5 times and the Chinese misunderstood the Swedes 4 times. Also, the Swedes could not understand the Chinese in 10 cases and the Chinese could not understand the Swedes in 13 cases. This may be because people coordinate with each other in the

interaction through, for example, adaptation and co-activation (Allwood & Lu, 2011) and that people may tend to have their understanding problems and difficulties occur closer to each other's in terms of frequency (i.e., the number of occurrences), time (i.e., when), and context (i.e., where). These assumptions could be further investigated in the future.

### 7.3.3 The role of prosody of a small pitch range in communication

Zuraidah and Knowles (2006) studied the prosody of the Malay language. They claimed that the prosody of a narrow pitch range (no clear definition of "narrow" in their paper) does not play any significant role in communication. However, the results from this thesis show that all the three types of understandings are most commonly expressed by vocal-verbal micro-feedback, which has a small pitch range (0–153 Hz as defined in this thesis). Also, another result from the thesis shows that only 22 out of 661 micro-feedback items have a large pitch range (307–459 Hz as defined earlier), and the large pitch range is not associated with any type of understanding at all (see Table 6.6). Thus, in spontaneous communication, the small pitch range plays a more significant role than the large pitch range. Although this does not reveal any distinct feature of sufficient understanding, misunderstanding, and non-understanding, it is something they have in common in terms of the pitch range.

### 7.3.4 The longer duration, the more useful information is for understanding?

Intuitively, we may assume that in interaction the longer the duration of the vocal-verbal response is, the more useful information for understanding it provides. Thus, short duration may be comparably insignificant in communication. However, the present study has found that both misunderstanding and non-understanding are associated with short duration (82–637 milliseconds as defined earlier), and sufficient understanding is associated with short and medium durations (638–1192 milliseconds as defined). All the three types of understandings have been found to be negatively associated with a long duration (1193–1748 milliseconds as defined) (see Table 6.8). This means that understanding in general is more associated with the vocal-verbal micro-feedback, which has a short duration, than vocal-verbal micro-feedback, which has a long duration. This is also a general feature of all the three types of understandings rather than anything distinctive. Thus, it is not always the case that "the bigger, the better". This is probably because the vocal-verbal micro-feedback expressions are normally unobtrusive and are mostly short in pronunciation and

word composition (i.e., duration). This is something that could be examined in future studies. Of course, full contributions and utterances help develop understanding and identify understanding. However, the length of the discourse context is not what has been measured and studied in this thesis, instead, it is the duration or length of the focused micro-feedback expressions.

### 7.3.5 Specific pitch contour patterns for sufficient understanding, misunderstanding, and non-understanding

The present study has found that sufficient understanding is negatively associated with the rising pitch contour but related to the same degree to the flat and falling pitch contours. Misunderstanding is associated with the falling pitch contour, and non-understanding is normally related to the rising pitch contour. These results are in line with Patel and Grigos' (2006) and Zuraidah and Knowles' (2006) findings, that a falling or a flat pitch contour is more frequently used than a rising one in statements and most sufficient understanding and misunderstanding cases are expressed in or related to statements rather than to other speech acts. Besides, a rising pitch contour is commonly used in questions, for example, asking for further clarification of some communicated and possibly perceived information, typically when non-understanding occurs.

## 7.4 Critical reflections and limitations of Study 1

As Cutrone (2010) suggested, the design and methods of any study or test to assess micro-feedback behaviour can vary and be highly dependent on the specific goals of the researchers as well as the technical support (e.g., research and analysis equipment and technologies) and the time they can get and afford when the study is conducted. The present study is limited in a few ways.

### 7.4.1 Communicative activity context

This particular context of Swedish–Chinese intercultural first encounters only provides us with a starting point for the analysis of micro-feedback in relation to understanding in spontaneous communication. Apparently, unacquainted people employ plenty of questions and answers to elicit and give micro-feedback in their conversations. However, they did not manifest many misunderstandings or non-understandings. Perhaps, first encounters are restricted to studying misunderstanding and non-understanding in the sense that unacquainted people normally do not want to reveal difficulties or problems in understanding in FTF spontaneous dyadic conversations. Thus, they probably try to minimise responses that show misunderstanding and non-

understanding. This first encounters' conversational context provides a considerable body of micro-feedback, but it does not provide as many cases of misunderstanding or non-understanding as it does of sufficient understanding.

It is also important to mention that micro-feedback behaviour is not only limited to contexts such as first encounters. It occurs in all types of human communications. Possibly, a larger group of dynamics and variables in terms of communication context could somewhat influence the micro-feedback behaviours and their associations with different types of understandings. However, it is not clear how much this influence would be and what form it would take. Consequently, further research with other communication activity contexts, cultures, and languages could shed light on these issues.

#### 7.4.2 An experimental situation

The present study of understanding in first encounters' conversations is based on an experimental situation rather than a naturalistic situation. The participants met in a lecture room. In contrast to Svennevig's study (1999), in which participants were going to attend a course together, here they did not foresee a long-term relationship for any joint social activity after their participation in this research project (unless, of course, some voluntary activity in private). They were simply given a task of getting acquainted with one another, preferably within eight minutes, in a face-to-face position standing up opposite to each other. Three cameras filmed the participants from different angles, and the participants were instructed not to move out of the cameras' capture areas.

In this situated interaction activity, on the one hand, the participants were aware of the activity rules, for example, what they were expected to do and what they could do. The participants developed sense-making and information sharing according to their knowledge of the activity type and in line with the activity expectation. On the other hand, the participants conformed to the rules of the current interaction activity, with a low level of awareness. The participants created their conversation jointly and interactively. From this perspective, in this study, the experimental situation may not vary much from a naturalistic situation in terms of the content of the conversation and the participants' interactivity. Especially, the use of micro-feedback has been identified primarily as of low consciousness and intentionality levels (see earlier chapters regarding literature and theoretical reviews). This setting does not invalidate the results of this particular study.

Because gesture and prosody are of great interest in the present study, capturing these elements in the interaction is important in the study design and material collection. An ethnomethodological study (of understanding in first encounters, e.g. in

a pub or business conference), which would have taken place without the study, may provide communication materials not provided by the researcher or the project. However, it may not be able to capture the interesting details of gesture and prosody in the real-time communication activity. Studies of gestures and prosodies usually use cameras and microphones (reasonably close to the participants). As long as there is a camera or microphone (or other artefacts for research purposes), a question that arises is how natural and ethnomethodological the study is.

### 7.4.3 Face-to-face communication

Although face-to-face (FTF) communication is historically seen as the basis of a theory of language, the basis of all human language behaviour, and the standard communication situation (Clark, 1996), the communication medium of FTF could have also constituted a restriction in the data in Study 1 on producing a greater number of understanding problems. Varying communication media can be explored in order to study micro-feedback and understanding in interaction.

### 7.4.4 Size of the data

Study 1 in the thesis was based on eight audio- and video-recorded FTF dyadic dialogues between four Swedish and four Chinese university students who were strangers, with two females and two males from each cultural group. Study 1 focuses particularly on the Swedish and the Chinese second language speakers of English and analyses their first encounter's conversations.

The size of the data is restricted in the sense that the cultural, language, gender, and educational backgrounds as well as the individual personalities and experiences could influence how they communicate in speech, gestures, and prosody.

First, the participants were not native English speakers, and Chinese and Swedish were their first languages. Although the participants were proficient users of the English lingua franca, it is uncertain how much the first language influences the second language usage. Therefore, how representative the results of the present study can be in relation to other people in other activities in the world is a question. Second, the participants were university students, studying at different universities and in different disciplines in Sweden. Their knowledge and experience of intercultural communication, in particular between Chinese and Swedish, and the frequency of using English may vary individually. Third, although general information on individual participants has been collected, for example, gender, age, education subject and level, general background information is not possible for any comparative analysis in the

present study because of the size of the data. Nevertheless, this background information can enhance the analyst's awareness of individual differences between the participants and provide the analyst with more opportunities to interpret, understand, and explain certain communicative actions and reactions, for example, the Chinese participant laughed in an "un-laughing" context (a situation where there is nothing that could be considered funny) when he did not understand the other participant and the Swedish participant did not understand the Chinese female participant's comment on her "beautiful eyes".

Consequently, the results presented in this thesis must be considered indicative. In line with what has been discussed earlier, an extended larger study involving, for example, a larger group of participants with more contextual variables could be conducted in the future.

## **7.5 Suggestions for future studies**

The subject of micro-feedback in relation to understanding can be studied with a larger group of dynamics and variables in terms of the communication context and participants. What has been studied is focused on an activity of a simple communication task of getting acquainted with one another. Thus, some types of communication with more complex communicative tasks than this could be of interest for future research. Casual conversation could provide important data for the analysis of micro-feedback and understanding (Verdonik, 2010). Thus, a larger set of casual dialogues (contexts) could possibly be added to the current research data. Exploring new activity types and communication contexts, such as educational and learning communication (e.g., teacher–student and student–student learning collaboration) as well as digital technology and computer mediated communication (e.g., Skype and FaceTime at the present time) involving both interpersonal and multimodal communication factors could be of interest.



## 8

## Results of reconceptualisation of understanding

Study 1, as presented earlier, did not generate many differentiated categories of understanding. Based on the results from Study 1, Study 2 expands the research by collecting new empirical data on a more complex communication activity with varying communication media.

Study 2 aims to conduct a further investigation of understanding in real-time communication, with a focus on how understanding problems are coped with by acquainted interlocutors in relation to not only micro-feedback but also other responsive interactions, primarily, meaning repair in an educational activity with collaborative learning tasks. It also aims to uncover similarities and differences in understanding between FTF and VMC.

Study 2 consists of three empirical analyses and addresses three research questions (RQ4–6). RQ4: What are understanding and understanding problems in social communication? RQ5: How are understanding problems detected, handled, and resolved in and through interaction? RQ6: What similarities and differences are there between face-to-face and video-mediated communication in the occurrence, detection, handling, and resolving of understanding and understanding problems?

The empirical material consists of 20 audio- and video-recorded interactions, ten FTF and ten VMC dyadic conversations, between ten Swedish and ten Chinese students who were 24 years old on average. The participants were acquaintances, and they were studying at universities in Sweden. They communicated in their common language, the English lingua franca. They were given a complex task of solving some learning assignments collaboratively.

Based on the empirical data, the study of understanding in real-time communication also carried out a more extended conceptual analysis of understanding compared to Study 1.

This chapter presents a theoretical attempt to reconceptualise understanding in real-time social interaction. The concept of understanding is examined by using an interactional approach based on theories of social communicative activity type, meaning and implicature, contextualisation, and relevance. Actions and reactions that are related to understanding, in particular, micro-feedback and meaning repair, are analytically focused on. In Chapter 8, one main research question will be investigated: what are understanding and understanding problems in social communication? The nuances of operationalising understanding will be presented.

## 8.1 Overview of the data in Study 2

By quantifying cases of varying understanding cases through micro-feedback and meaning repair in interaction (for more discussion see Section 2.12), an overview of the data on understanding in Study 2 can be obtained.

*Table 8.1.* Understanding occurrences in Study 2. (Comm. Situation = communication situation; Time = hour:minute:second; Word/Utterance = number of words and utterances; Suff. U = sufficient understanding; In suff. U = insufficient understanding; Mis U = misunderstanding; Partial U = partial understanding; Non U = non-understanding)

Comm. Situation	Length			Suff. U	In suff. U/Understanding problems			
	Time	Word	Utterance		Mis U	Partial U	Non U	Total
VMC	1:37:31	10815	1446	1026	14	40	33	87
FTF	1:31:57	9854	1573	1025	7	29	30	66
Entire data Total	3:09:28	20669	3019	2051	21	69	63	153

The empirical data in Study 2 contain, as shown in Table 8.1, 2051 cases of sufficient understanding and 153 cases of understanding problems which consist of 21 cases of

misunderstanding, 69 cases of partial understanding, and 63 cases of non-understanding.

As shown in the table, there are 2204 (i.e., 2051+153) cases of understanding and understanding problems that are often associated with micro-feedback expressions, and there are 3019 utterances. This means that there are 815 (i.e., 3019–2204) utterances that are irrelevant to the analyses. For example, some utterances are intended to initiate the interaction such as *now I guess we can start*; some for social greetings such as *hello*; some are for exchanging situational emotions and attitudes, for instance, *this is so weird it feels like I'm in classes*; some are for stressing the project task such as *so pick up three things that are similar*; some are for linking and altering sequential topics such as *anything else, we need to pick up three points, now we have one already, so what is the next, I was expecting a eh a discussion ah so if you can start because you picked things*; and some are for closing the interaction, for instance, *so we finished, shall we call the researcher* (all taken from the empirical data in Study 2). These are typically utterances that do not belong to the substantial matters (topics) of the task (cultural similarities and differences between Sweden and China), and they are therefore not included in the analysis of understanding and understanding problems in this thesis. Only the utterances and gestural contributions that contain responsive actions, either through micro-feedback or other communication behaviours primarily meaning repair related to some information earlier presented, are taken into account.

## **8.2 Quantifying understanding by primary means of micro-feedback**

As presented earlier in Section 2.12, as regards the quantitative occurrences of understanding in the data analysis, efforts are made to code and count micro-feedback units and other related responsive actions, primarily, meaning repair, that are signalling or related to different forms of understanding. This is primarily because that micro-feedback has functions such as acknowledgement of various understandings of the intended content of the interlocutor's prior utterances and gestural contributions (see more in Sections 1.2 and 2.10).

At present, this quantitative method for approaching understanding may not be optimal, but it serves the purpose of studying and evaluating understanding from a quantitative perspective in this thesis. More discussions about it can be found in Section 2.12.

### 8.3 Analytical result of reconceptualisation of understanding

Figure 8.1 presents the analytical results of this chapter, that is, the theoretical classification of understanding in social interaction. Understanding can be categorised into sufficient understanding and insufficient understanding (or understanding problems), with the latter including misunderstanding, non-understanding, and partial understanding. More subtle degrees of each form of understanding are not taken into account in this thesis.

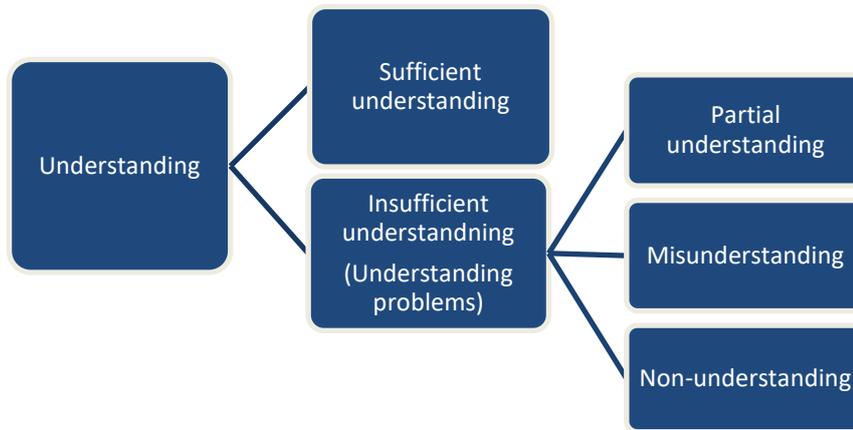


Figure 8.1. Theoretical classification of understanding.

Figure 8.2 illustrates the relation between understanding and communication, that is, theorising understanding in communication practice. Sufficient understanding leads to successful communication, whereas insufficient understanding results in miscommunication.

As found in the empirical data, sometimes understanding in interaction corresponds to what is intended and anticipated, thus becoming sufficient understanding and leading to successful communication (see Figures 8.1 and 8.2). At other times, understanding deviates from what is intended and anticipated, thus becoming insufficient understanding and resulting in miscommunication (see Figures 8.1 and 8.2 too).

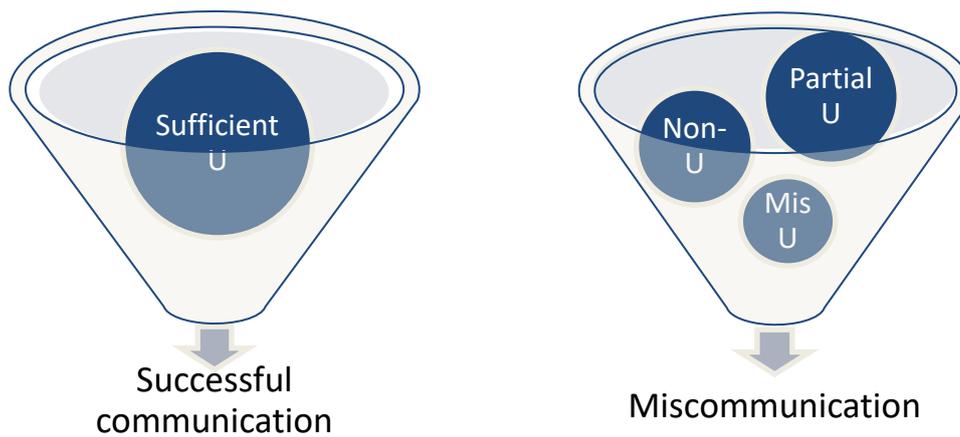


Figure 8.2. Understanding and understanding problems in relation to successful communication and miscommunication.

Theoretically, sufficient understanding can consist of more than intended and anticipated meanings. For example, the speaker may learn something new, or realise that his prior intentions were misguided. However, first, such cases were not obvious for the analyst in the empirical data studied and are therefore not analysed. Second, in such cases, the sufficient understanding that is achieved eventually after a number of sequential information sharings and understandings is a very global understanding (of a long sequence of context). Instead, the local understanding of the listener in each discourse exchange is the focus of this thesis (see more discussions in Section 2.4, examples are presented below).

### 8.3.1 Sufficient understanding

As discussed earlier in Chapter 2, understanding usually is unfinalisable and there is no complete or full understanding (except possibly of requests and supply of plain information such as name, address, age, etc.). In this sense, understandings are usually partially shared understandings (see earlier for more philosophical discussions).

Sufficient understanding refers to the partially shared understanding that is sufficient to serve the current practical purposes of information sharing, sense-making, and continuing communication, no matter how much is partially but sufficiently shared. Also, sufficient understanding is usually exhibited in the speech acts of declaring and persuading. In sufficient understanding, the information presented is understood in a way that is correct for current purposes in relation to what is intended and anticipated, no matter how much is correct. The interlocutors are content with

the understanding of one another and feel good enough to proceed further in their communication. Sufficient understanding leads to successful communication.

**Excerpt 1: Sufficient understanding example: extracted from D16d(150309).v, pages 1–2**

- 1 Ou: <eh> we could talk about
- 2 Patrik: the hierarchy
- 3 Ou: yes
- 4 Patrik: that in china that it is more common that the leader [eh]
- 5 Ou: [decide] [the situation yes]
- 6 Patrik: [decides] yeah m
- 7 Ou: and make the decision
- 8 Patrik: yes (combined with head up-nods signalling agreement)
- 9 Ou: so and in sweden it's more like everybody should agree this decision
- 10 Patrik: m (combined with head up-nods, making confirmation)
- 11 Ou: okay that's one point (in combination with chuckle)
- 12 Patrik: yeah okay

As found in the empirical data, most often people achieve sufficient understanding and their interaction continues as smoothly as anticipated. Excerpt 1 presents a number of sufficient understanding cases between participants Ou and Patrik. In Line 2, Patrik suggested talking about hierarchy differences between China and Sweden. Ou showed her sufficient understanding and agreement by replying with a micro-feedback *yes* in Line 3, which is correct as anticipated by Patrik. This (Line 3) is the first case of sufficient understanding in this interaction sequence.

Then, the interaction continued smoothly as intended, and Patrik said that *in china it is more common that the leader decides* in Lines 4 and 6. Apparently, in Line 5 in-between 4 and 6, by saying *decide the situation yes*, Ou had already sufficiently understood what Patrik intended to say even before Patrik had finished his complete utterance by Line 6. The utterance in Line 5 is the second case of sufficient understanding in this sequence.

In Line 6, Patrik heard what Ou said in Line 5 and Patrik acknowledged his understanding and agreement by saying *yeah m*. The utterance in Line 6 is the third case of sufficient understanding in the interaction sequence.

Then, Ou complemented her earlier utterance in Line 5 with another utterance *and make the decision* in Line 7. Patrik sufficiently understood it as intended and anticipated, and then signalled it with a micro-feedback *yes* in combination with head up-nods in Line 8, which serve the current practical purpose of sharing information and looking for solutions to their task. The fourth case of sufficient understanding occurs in Line 8.

Following this, in Line 9, Ou related the topic of hierarchy to Sweden and said *in sweden it's more like everybody should agree this decision*, which changes the focus

from China to Sweden and asks for the Swede's confirmation. In Line 10, Patrik responded with micro-feedback *m* accompanied by head up-nods, which signal sufficient understanding and confirmation. Line 10 is counted as the fifth case of sufficient understanding in this discourse sequence.

Then, Ou said in Line 11 that *okay that's one point*, which signals sufficient understanding of what Patrik had said in Line 10. This (Line 11) is the sixth case of sufficient understanding in this sequence.

What Ou said in Line 11 was later on sufficiently understood and agreed on by Patrik, saying *yeah okay* in Line 12, which is counted as the seventh case of sufficient understanding. These seven cases are counted by primary means of micro-feedback (more discussions have been presented in Sections 2.12 and 8.2).

Sufficient understanding not only serves the current practical purpose of sharing information and maintaining interaction, but also makes sense of the exchanged information in a correct way as intended and anticipated. With these anticipated sufficient understanding interactions (actions and reactions to one another), both the participants felt content with the proceedings and their interaction proceeded smoothly and successfully as intended.

**Excerpt 2: Sufficient understanding example: extracted from D17d(150310).v, page 2**

- 1 Rikard: eh: I think power distance is probably the biggest one
- 2 Qing: yeah
- 3 Rikard: eh: in sweden
- 4 Qing: head nod
- 5 Rikard: it's I never called my boss boss
- 6 Qing: yeah
- 7 Rikard: even when I worked in the army
- 8 Qing: yeah
- 9 Rikard: I don't I rarely called my superior officers their rank I just called their name
- 10 Qing: yeah yes yes chuckle yeah
- 11 Rikard: anything else
- 12 Qing: <em> and I think about eh how people use salary how people use the money ...

Excerpt 2 presents another sequence of utterances full of sufficient understandings. Here, participants Rikard and Qing both agreed that comparing to other cultural issues (mentioned in the reading material), “power distance” is the biggest difference between China and Sweden. Rikard was primarily giving examples of his own experience of “low power distance in Sweden” (see Lines 3, 5, 7, and 9), and Qing was primarily listening and responding with micro-feedback signalling sufficient perception and understanding of the information presented (see Lines 2, 4, 6, 8, and 10). In Line 11, Rikard initiated a change to a local topic, and sequentially Qing accepted it

and came up with a suggestion of *how people use salary and money* in Line 12. Apparently, their interaction continued smoothly and successfully as intended and anticipated, and both were content with what was happening. This sequence of interaction appears to be full of sufficient understandings, which are formed and developed in a correct way as intended and anticipated by the participants.

### 8.3.2 Insufficient understanding: understanding problem

It is not uncommon that people cannot achieve sufficient understanding as they would ideally wish. Instead, different understanding problems may occur all the time throughout the interaction.

As opposite to sufficient understanding, insufficient understanding occurs when there is an understanding problem, the interaction may or may not serve the current practical purposes of information sharing, sense-making, and continuing communication, and the interaction does not proceed as intended and anticipated. Except for sufficient understanding, all the other forms of understanding, including misunderstanding, partial understanding, and non-understanding, are regarded as insufficient understanding and are thus treated as understanding problems in this thesis. Insufficient understanding or understanding problems can be developed into sufficient understanding but not always, and usually leads to miscommunication.

#### Misunderstanding

Misunderstanding is one form of insufficient understanding or understanding problem. Misunderstanding only occurs when the information is understood in an incorrect way that deviates from what is intended or anticipated, although it can perhaps serve the current practical purposes of exchanging information and carrying on the interaction. Misunderstanding occurs usually in speech acts in the form of declaring and persuading statements, similar to sufficient understanding, which makes it difficult to differentiate. Misunderstanding can sometimes serve the purpose of achieving sufficient understanding because it may give rise to repair. However, it can sometimes be persistent or cause additional misunderstandings and therefore not lead to anything close to sufficient understanding. Misunderstanding is often not detected or detectable by the interlocutors. The interlocutors continue the interaction without being aware of the occurrence of misunderstanding.

**Excerpt 3: Misunderstanding example: extracted from D2s(150121).f, pages 5–6 (Note: the underlined utterance represents the occurrence of the target item in Study 2.)**

- 1 Ann: another thing ... eh: similar  
...  
2 Bengt: and then china you don't have fika

- 3 Ann: yeah we don't have  
 4 Bengt: well it's | but you eat a lot  
 5 Ann: laugh (misunderstanding, caused by different concepts of "a lot", embarrassment/amusement)  
 6 Bengt: people like people they they gather to eat you know  
 7 Ann: yeah yeah yeah yeah like to eat  
 8 Bengt: yeah  
 9 Ann: actually after the work  
 10 Bengt: yeah  
 11 Ann: eh: [till the] till the night  
 12 Bengt: [that's not so com]  
 13 Ann: the late night  
 14 Bengt: yeah aha  
 15 Ann: ah: I think chinese people like to ... drink wine  
 16 Bengt: < drink wine >

Excerpt 3 shows that the Chinese participant Ann and the Swedish participant Bengt were looking for another similarity between China and Sweden. First Bengt mentioned *china you don't have fika* in Line 2, and Ann agreed that *we don't have* in Line 3. Then, Bengt said that *but you eat a lot* in Line 4, which was apparently misunderstood by Ann as that Chinese people eat large amounts of food. This is likely caused by different concepts of the English phrase "a lot". So Ann did not say anything but only gave a laughing response, which may signal embarrassment and/or amusement (see Line 5). Apparently, from Ann's reaction, Bengt immediately realised that his earlier utterance was not clear enough, so he initiated a meaning repair in the next utterance (Line 6) *people like people they they gather to eat you know*. There, Bengt expressed more clearly what he intended was likely that Chinese people eat a lot together, and he even added *you know* in the end of his utterance to stress his intention and anticipation, likely that "Chinese people gather a lot to eat together is what I actually meant, you know, and I did not mean that you eat a large amount".

This could be interpreted differently by different people, but this is the analyst's immediate interpretation. First, Bengt and Ann are classmates and friends in private life, and they are formally participating in a research project and collaborating to complete the project task within the time set and according to the content requirements, so it does not make much sense that Bengt was teasing Ann with "Chinese people eat too much" in this context. Second, in the earlier context sequence, Bengt talked about people in Sweden not wanting to socialise much but rather keeping some distance from each other, so it is very natural that he turned to talk about China in that "Chinese people stay a lot together and eat a lot together". In fact, it is very rare to see a single Chinese person eating in a restaurant, especially in a place outside China. It would not be too much to say that it is absolutely impossible for a Chinese person to go to after-work by him- or herself, except, of course, in very special situations. It is Chinese culture, a culture of being together and socialising at certain

events and occasions such as dining, drinking, dancing, karaokeing, and pubbing. Therefore, from an analytical perspective, Ann's laughter in Line 5 is annotated as an action of misunderstanding, and Bengt's follow-up reaction in Line 6 is annotated as his own initiative of meaning repair.

Again, it is not certain that Bengt's utterance in Line 6 is a meaning repair, nor how Ann has understood Bengt's meaning repair afterwards. All that can be seen is that after Bengt's repair in Line 6, it is good enough for Bengt and Ann to carry on the interaction and continue their communication task. It seems that Bengt's repair worked and Ann understood sufficiently. As uttered in the following Lines 7, 9, and 11, Ann agreed that Chinese people like eating (together) after they finish work and usually eat until late at night. This understanding problem occurred probably because Bengt did not communicate or provide the complete information he had intended for the discourse exchange. That is, this type of understanding problem is caused by incomplete information from one interlocutor to the other.

**Excerpt 4: Misunderstanding example: extracted from D4s(150123).v, pages 5–7 (in two occurrences of misunderstanding)**

- 1 Chun: yeah maybe uncertainty avoidance
- 2 Daniel: where where where are you reading
- 3 Chun: eh: the one two the fourth page
- 4 Daniel: first page yeah (misunderstanding caused by misperception of the English word "fourth")
- 5 Chun: fourth
- 6 Daniel: firth okay (misunderstanding, still because of misperception of the English word "fourth")
- 7 Chun: one two three the fourth paragraph uncertainty avoiding
- 8 Daniel: em: oh (showing emotions of surprise/disagreement)
- 9 Chun: because I think all the people want to get a clear information eh we try to avoid the
- 10 uncertainty so I think it's eh similar
- 11 Daniel: avoid the uncertainty I'm not quite sure of how you mean (partial understanding, because of
- 12 wanting more clarification and explanation)
- 13 Chun: eh: (shows hesitation)
- 14 Daniel: could you explain again (asks for more clarification)
- 15 Chun: I don't know if I misunderstand this but I think it's refer to when you get some information
- 16 it's not so clear you feel confused and you want to avoid this kind of things
- 17 Daniel: head nods

In Excerpt 4, there are two cases of misunderstanding problems. One occurred in Line 4 and the other in Line 6. Both are caused by Daniel's misperception of the English word *fourth*, and both are in the end resolved by Chun with her clarification in Line 7 *one two three the fourth paragraph uncertainty avoiding*. Apparently, the two cases of misunderstanding presented in Excerpt 4 were well observed by participant Chun at the same time as they occurred. They were also immediately repaired and corrected by Chun in the interaction sequence. The two cases of misunderstanding

were in the end resolved and actually promoted and contributed to sufficient understanding, although this is not always the case in interaction. In Chapter 9, a number of examples, where misunderstanding is neither observed nor resolved and sometimes even leads to further misunderstandings, will be presented.

### Non-understanding

Non-understanding is another form of insufficient understanding or understanding problem. Non-understanding occurs when the information presented is not understood at all, for reasons such as lack of access to the presented information itself or some background knowledge of relevance. Non-understanding cannot serve the current communication purposes of sharing or making sense of the information presented. It does not correspond to what is intended and anticipated. Normally non-understanding is manifested in the speech act of questions, seeking more clarification. Sufficient understanding can be usually achieved in and through further interaction.

#### **Excerpt 5: Non-understanding example: extracted from D13s(150220e).v, pages 1–2**

- 1 Min: I guess we start
- 2 Niklas: yeah
- 3 Min: yeah so we have to agree if I understand right we have to agree on three points we think similar
- 4 Niklas: yeah
- 5 Min: and then one of those is the most important [like the]
- 6 Niklas: [yeah]
- 7 Min: okay I've picked several < smile >
- 8 Niklas: really
- 9 Min: [yeah I picked five]
- 10 Niklas: [I haven't picked any] one yet I was thinking we could have discussion I guess now there is a
- 11 eh difference
- 12 Min: m (eliciting more clarification with a rising prosody, this is a non-understanding case)
- 13 Niklas: I guess that's for the difference I haven't picked anything
- 14 Min: < chuckle > okay
- 15 Niklas: I was expecting a eh a discussion ah so if you can start because you picked [things]
- 16 Min: [but] maybe you disagree like sometimes I don't know yeah the first one I think like the most
- 17 similar one could be on page 43 m yeah like the yeah the page 43 is the specific versus diffuse said
- 18 that in the specific cultures individuals are direct clear bland and to the point examine the facts and
- 19 in the diffused cultures individuals are more indirect and humble
- 20 Niklas: yes

Excerpt 5 shows that Min and Niklas were discussing the project task of picking out three similarities between Swedish and Chinese cultures and also pointing out the most important one. Min said *okay I've picked several* in Line 7. Then, Niklas said *I haven't picked any one yet I was thinking we could have discussion I guess now there is a eh difference* in Lines 10 to 11. Min became confused by Niklas' response. Min

did not understand what Niklas meant, and she became uncertain about the conversation. So, Min uttered *m* in a rising prosody in Line 12, with the intention of seeking further clarification and resolving the understanding problem. Then, Niklas explained in the following utterance (Line 13) that *I guess that's for the difference I haven't picked anything*. This enabled Min to understand sufficiently what Niklas meant earlier, that Min had already picked out several similarities in order to solve the task but Niklas had not picked out any, so Niklas concluded that this was indeed a difference. Then, Min chuckled and said *okay* in her response (Line 14), which signals that Min has so far achieved a sufficient understanding of Niklas' humorous way of saying "difference" versus "similarity" in the prior utterances. Apparently, both Niklas and Min were content with the current proceedings and the interaction continued as anticipated. After Line 15 (see Lines 15 to 20), the participants even started a new topic right afterwards. Therefore, from an analytical perspective, it is interpreted as that Niklas and Min had sufficiently understood one another by Line 14, and the case of a non-understanding problem was resolved. This non-understanding problem is probably caused by different senses of humour between the Chinese and Swedish participants. Obviously, the Swede chose to play a language game by saying "difference" in a humorous way as a contrast to "similarity", but the Chinese did not get it simultaneously as anticipated.

### Partial understanding

Partial understanding<sup>52</sup> is another form of insufficient understanding or understanding problem. Partial understanding takes place when the interlocutor cannot sufficiently, but only partially make sense or share the meaning of the information presented and the interaction does not continue as intended and anticipated. Partial understanding is usually revealed through an eliciting utterance of a suggestion or question, seeking confirmation, clarification, explanation, or specification about what has been immediately perceived and understood. Partial understanding is a developing stage towards achieving sufficient understanding, and it usually succeeds eventually.

Excerpt 6 demonstrates two cases of partial understanding. The first one occurred in Line 4 when Bengt was not completely sure if Ann was talking about wine. There, Bengt initiated an eliciting question by repeating the perceived information to ask for confirmation. This partial understanding problem is probably caused by different concepts of the English word *wine*. The Swedish cultural conception of "wine" is that it was originally French and that there was no Swedish wine (see

---

<sup>52</sup> Note that this is not referring to the fact that understandings are arguably "partial" in that they are not complete in an abstract theoretical sense (see Chapter 2, primarily Section 2.7).

Bengt's later explanations in Lines 9 and 11). This does not correspond to Ann's conception of "wine" in the Chinese culture where almost all types of liquor are called "wine".

**Excerpt 6: Partial understanding example: extracted from D2s(150121).f, pages 7–8 (in two occurrences of a partial understanding problem)<sup>53</sup>**

- 1 Ann: [but swedish] people like to drink wine because of the cold weather or like | or the socialising
- 2 Bengt: social not so much about the cold weather
- 3 Ann: < chuckle > because I think the swedish wine is so heavy maybe
- 4 Bengt: okay wine (with prosody stressing/asking for confirmation, partial understanding)
- 5 Ann: why why eh: [wine wine] (partial understanding, seeking more clarification and explanation)
- 6 Bengt: [wine]
- 7 Ann: yeah
- 8 Bengt: okay (Ann laugh)
- 9 Bengt: but it's | it's French wine from the beginning you know
- 10 Ann: ehm
- 11 Bengt: and eh: it ended up at eh:
- 12 Ann: so maybe we say it's the similar like to socialising [you mean]
- 13 Bengt: socialise in drinking
- 14 Ann: yeah
- 15 Bengt: let's call it like that

Sequentially, in Line 5, the second instance of partial understanding occurred when Ann did not sufficiently understand what Bengt intended. So she asked *why why eh: wine wine* in order to seek more clarification and explanation. Then, Bengt made more efforts in the following utterances to clarify what he had meant earlier. The sequence shows that Bengt was focusing on clarifying the western concept of "wine" while Ann was mainly focusing on addressing the point of socialising when drinking alcohol. Although it is not easy to judge how much Ann has eventually understood the historical cultural background of "wine", in one way or another Bengt and Ann

---

<sup>53</sup> The earlier context of this excerpt is:

Ann: <ah:> I think chinese people like to

Bengt: have after work

Ann: < | > drink wine (chuckle, showing disagreement)

Bengt: < drink wine > (with head nod)

Ann: I think the swedish people like to drink wine but | yeah chinese people like to drink wine because they want to | make friends or

Bengt: socialise socialise

Ann: < yeah > socialise

Bengt: so drinking for socialise

Ann: < yeah > (with head nods)

Bengt: is that common between sweden and china then

Ann: < yeah > (chuckle, showing agreement/joyfulness)

Bengt: and common in most of [the world]

had reached a sufficiently shared understanding and their interaction continued and developed successfully.

### 8.3.3 Distinction between different forms of understanding

*Sufficient understanding* differs from insufficient understanding in that sufficient understanding is good enough for the interlocutors to carry on the communication, the information is considered to be understood in a way that is correct for current purposes in relation to what is intended and anticipated, and the interlocutors feel satisfied that they have achieved the current purpose of communication. Sufficient understanding underpins smooth and successful communication without eliciting questions or showing uncertainties or difficulties when processing the information communicated. Insufficient understanding is the opposite.

*Partial understanding* is one type of insufficient understanding. In a case of partial understanding, one understands neither sufficiently nor does one understand anything at all. It is a developing understanding process that always progresses further along the interaction and usually leads to sufficient understanding in the end, either globally or locally (i.e., the context is either a long sequence or a relatively short one). Partial understanding is often signalled in questions such as *do you mean this...*, *is it this...*, and the like. Interaction involving partial understanding does not flow as smoothly as that involving sufficient understanding. It usually involves meaning repairs and negotiations that resolve the uncertainties, doubts, and questions about the information communicated and helps to achieve sufficient understanding eventually.

Many researchers tend to merge partial understanding with misunderstanding or vice versa (e.g., Zaefferer, 1977; Hirst et al., 1994; Weigand, 1999; Verdonik, 2010; Mustajoki, 2012). However, partial understanding and misunderstanding are regarded as separate types of understanding problems in this thesis. It is found that partial understanding not only involves sense-making and information sharing and serves the current practical purposes but also develops understanding in a correct way in relation to what is intended and anticipated, which usually leads to sufficient understanding eventually. This is normally very different from misunderstanding.

*Misunderstanding* does involve sense-making and information sharing and does serve current practical purposes, although in an incorrect way in relation to what is intended or anticipated. Interaction involving misunderstanding can continue as smoothly as that of sufficient understanding. Sometimes misunderstanding may be detected and corrected, but very often one misunderstanding instance results in further misunderstanding instances that are not detected at all throughout the conversation. Although interlocutors may carry on the interaction without being aware of

any misunderstanding occurrence, very often misunderstanding does not lead to sufficient understanding or even anywhere close.

*Non-understanding* differs from misunderstanding in the sense that it does not involve any achieved sense-making of the information communicated at that moment. Non-understanding does not serve the current practical purposes of sharing and making sense of the information presented. It does not correspond to what is intended and anticipated. However, by eliciting questions such as typically *sorry*, *what did you say*, *please pardon*, *what do you mean*, *can you please say that again* and the like, non-understanding occurrences usually help to develop sufficient understanding. Non-understanding is like partial understanding, in that they are both detectable and resolvable and also they are usually developed in time into another form of understanding. Similar to partial understanding, interaction containing non-understanding does not flow as smoothly as interaction containing sufficient understanding.

### 8.3.4 Possible reasons for occurrences of understanding problems

Because one instance of an understanding problem can be caused by several complex reasons, it is not practical or possible to address them all. Therefore, among several possible reasons, the one that is more likely than others (from an analytical perspective) was coded and analysed for each occurrence of understanding problem. This means that each understanding problem occurrence has one coded reason in this study. By using an inductive method, possible reasons for the occurrences of the understanding problems studied have been identified and listed in Table 8.2.

*Table 8.2.* The likely causes of understanding problems in the empirical data (ordered by frequency, starting with the most frequent category).

<b>Reasons for understanding problems</b>	<b>Mis u</b>	<b>Part u</b>	<b>Non-u</b>	<b>Total</b>
New to knowledge and experience, new topic, or new terminology	1	60	46	107
Diff. concepts of same Eng. word or perspectives of same concept	12	3	1	16
Limit of language proficiency in vocabulary	0	5	5	10
Limit of language proficiency in pronunciation (mispronunciation)	1	0	5	6
Faded memory	0	1	5	6
Further misunderstanding caused by earlier misunderstanding	4	0	0	4
Misperception of an English word or sentence	3	0	0	3
Different senses of humour	0	0	1	1
<b>Total</b>	<b>21</b>	<b>69</b>	<b>63</b>	<b>153</b>

As presented in the table above, new to knowledge and experience, new topic, or new terminology is the most frequent cause of understanding problems in the data. For example, in one case, participant Kang asked participant Louise *have you read this Hofstede four dimensions*. Louise did not understand this new content and asked *which one* and *which page is it* to seek more clarification and details. In another example, participant Chun suggested a new topic *relation to the nature* as another solution to their task. Participant Daniel partially understood it and repeated *relation to the nature* in an eliciting tone to ask for further explanation. Reasons such as different concepts of the same English word or different perspective of the same concept and limit of language proficiency in vocabulary and pronunciation are also frequent. They are discussed to varying extents in the excerpts in this thesis.

Furthermore, the data have shown some association between the likely causes and the different types of understanding problems. First, compared to the other two types of understanding problems, misunderstanding is more frequently caused by reason of different concepts of the same English word or different perspectives of the same concept. This suggests that linguistic and conceptual differences are the main reasons for misunderstanding occurrences. Also, only misunderstanding is associated with reasons for further misunderstanding caused by earlier misunderstanding and misperception of an English word or sentence. Second, both partial understanding and non-understanding are more commonly associated than misunderstanding with the causes of new to knowledge and experience, new topic, or new terminology and limit of language proficiency in vocabulary. This suggests that knowledge, experience, and language competence are the primary causes of partial understanding and non-understanding. Third, compared to partial understanding and misunderstanding, non-understanding occurred more frequently when associated with the limit of language proficiency in pronunciation (i.e., mispronunciation) and faded memory (which could be caused by disagreement at the beginning and inability to remember how one interlocutor argued and the other interlocutor became persuaded). Only non-understanding is caused by different senses of humour. Since exploring the likely reasons for the occurrences of understanding problems is not the focus of the study, this issue will not be addressed further.

## 8.4 Discussion

Based on the new empirical data in Study 2, sufficient understanding and insufficient understanding or understanding problems, which include misunderstanding, non-understanding, and partial understanding, have been identified. This analytical result is along the lines of the earlier theories of *sufficient understanding* in Garfinkel

(1967), Bakhtin (1986), Taylor (1992), Linell (2009), and Lindwall and Lymer (2011), *partial understanding* in Zaefferer (1977) and Allwood (1986), *misunderstanding* in Zaefferer (1977), Allwood (1986), and Weigand (1999), and *non-understanding* in Zaefferer (1977) and Weigand (1999), although the definitions may be somewhat different.

#### 8.4.1 Evolving from Study 1 to Study 2: methodological approach to studying misunderstanding

As discussed in Study 1, not all the understanding cases are communicated through micro-feedback but all the micro-feedback expressions are in relation to some kind of understanding. To be more specific, Study 1 has found that non-understanding is always expressed through micro-feedback, whereas sufficient understanding is most often but not always and misunderstanding is usually not expressed through micro-feedback. On the contrary, very often misunderstanding is not signalled or expressed through micro-feedback. Rather, misunderstanding is revealed through other responsive communicative actions vocal-verbally, gesturally, and prosodically in the contexts that are of relevance. Study 1 has proved that contextualisation and relevance theories help to identify and analyse misunderstanding in human communication.

In Study 2, micro-feedback as one possible signal of understanding can help to identify most of the sufficient understanding cases, all the non-understanding cases, and some misunderstanding cases. This may benefit my research purposes in Study 2 in this thesis. However, as observed from Study 1, misunderstanding does not occur as often as sufficient understanding or non-understanding; therefore, in Study 2, the coding of micro-feedback in relation to understanding was expanded to include all types of understanding problems that can be found in the data. With this methodological expansion, most of the sufficient understanding cases, all the non-understanding and misunderstanding cases, and all the remaining possible types of understanding problems that may emerge from the new data are included in the analysis.

From my point of view, two thousand sufficient understanding cases are enough for analysis, and what is more critical for the analysis is an enlarged volume of data on understanding problems. In Study 2, the earlier coding of micro-feedback in relation to understanding used in Study 1 was retained and evolved in order to code all the analytically observed understanding problems. The purpose was to obtain a richer data on understanding (problems in particular) in interaction and to approach it in a more sophisticated way.

#### 8.4.2 New finding in Study 2: theoretical framework of partial understanding

In this thesis, the general criterion for classifying understanding is the same as in both Study 1 and Study 2, relying on whether the information is understood sufficiently and correctly in relation to what is required to continue the interaction and what can be inferred about the interlocutor's intention and anticipation.

In Study 1, partial understanding was not found in the empirical data of first encounter studied, thus the underlying theoretical framework used there focused on sufficient understanding, misunderstanding, and non-understanding. In Study 2, with the new empirical data of educational task-solving collaboration, partial understanding emerged, and was identified as a new category of understanding. As a nuance, the emergence of partial understanding in Study 2 rather than in Study 1 in fact constitutes the main difference in the theoretical framework between the two studies in this thesis.

As a continuing project from Study 1, Study 2 was designed as an evolvement of Study 1 in terms of both theoretical foundation and methodological approach. It is to some extent a research progress of Study 1, and it provides more insights into the understanding of understanding problems of social communication. This new finding of partial understanding seems to be in accordance with the theories of social communicative activity type (see more in the following section).

#### 8.4.3 Social communicative activity types in Study 1 and Study 2: simple and complex contexts

As described earlier, this thesis consists of two parts. Study 1 explores micro-feedback in relation to understanding issues in a spontaneous communication activity in first encounters, with a simple activity task of getting acquainted. Based on the results from Study 1, Study 2 expands the research and attempts to carry out a further investigation of how understanding problems are coped with by acquainted interlocutors in relation to not only micro-feedback but also meaning repair in an educational activity, with the more complex task of solving some learning assignments collaboratively.

Interaction content, language use, communicative genre, and mutual understanding have been found to vary between these two communicative activities. In Study 1, the participants tend to concentrate on identifying and sharing personal information such as who they are, what their names are, how old they are, where they are from, what they are studying, why come to this city, what interests them, where they live, what they have done before, what they plan for the future, whether they

have family, what they like or dislike about Sweden, how they joined the research project, and so forth. Several of these topics can be addressed quite efficiently by means of simple questions and answers in exchange of plain information transfer. The language used is quite simple and interpersonal in its vocabulary. The communicative genre is fairly casual, autonomous, and everyday. Mutual understanding is the rule: overwhelmingly sufficient understanding, occasionally non-understanding, and rarely misunderstanding. Partial understanding was not found in Study 1.

In Study 2, however, the participants have to cooperate to solve two learning assignments in face-to-face and video-mediated communication situations with the help of reading material provided by the project. The participants first read the provided material and tend to look for solutions themselves, then they share with peers, next discuss, negotiate, reason, persuade, and agree or disagree with one another and finally, form, complete, and produce joint task solutions. This particular activity type is normative in nature. The language used is relatively extensive, terminological, and technical in its vocabulary. The communicative genre is comparatively more formal, goal-driven, and institutional. More varieties of understanding have been found in the data: most often, sufficient understanding; second most frequently, partial understanding; sometimes non-understanding; occasionally misunderstanding.

This result is consistent with Sins et al.'s (2011) and Hancock and Dunham's (2001) studies in that communicative understanding behaviours vary between simple and complex communication tasks. Compared to first encounter's task of getting to know each other, collaborative learning task-solving requires a more intensive discourse exchange, has higher demands on understanding, and reveals more varieties of understanding problems.

This variance between Study 1 and 2 is along the lines of theories of Wittgenstein's (1958) language games, Vygotsky (1962) and Rommetveit's (1974) language and its social interactive functions, Grice' (1975) meaning and implicature, Levinson's (1979) activity type, Allwood's (2013) activity-based communication analysis, Gumperz' (1982) relevant aspects of activity type, and Linell's (2010) communicative activity type. When we engage in interaction, we are always in a social activity. Language can be only understood in social activity. Social activity affects language use, for example, communication and mutual understanding. An interactional approach of a social communicative activity type provides a link between situated interactions and societal premises for such interactions.

## 8.5 Conclusion of Chapter 8

With the aim of further studying understanding in real-time communication, this chapter, departing from Study 1 and entering Study 2, has made an attempt to re-conceptualise understanding in new empirical data. Understanding is categorised into sufficient understanding and insufficient understanding, that is, understanding problem, which includes partial understanding, misunderstanding, and non-understanding.

Sufficient understanding leads to successful communication and insufficient understanding leads to miscommunication and unsuccessful communication. The criterion lies in whether the information is understood sufficiently well and correctly in relation to what is required to continue the interaction and what can be inferred about the interlocutor's intention and anticipation.

In the empirical data studied, sufficient understanding has been identified as occurring most frequently, followed partial understanding, non-understanding, and finally misunderstanding, which occurs with the lowest frequency.

Sufficient understanding occurs when it is enough to serve the current practical purposes of information sharing and sense-making, no matter how partially the understanding is shared. It is usually exhibited in the speech acts of declaring and persuading. The information presented is understood in a way that is correct for current purposes in relation to what is intended and anticipated, no matter how much is correct. Interlocutors are content with understanding one another and feel good enough to proceed further.

Partial understanding is one form of insufficient understanding or understanding problem in this thesis. Partial understanding occurs when one cannot sufficiently but only partially make sense or share the meaning of the information presented as it is intended or anticipated. Partial understanding is usually revealed by eliciting suggestions or questions, seeking further confirmation, clarification, explanation, or specification. It is a developing stage towards achieving sufficient understanding, and it is usually successful eventually. Partial understanding is only observed in task-solving interactions in Study 2 and not in first encounters in Study 1.

Misunderstanding is also a form of insufficient understanding or understanding problem. It only occurs when the information is understood in an incorrect way or deviates from what is intended or anticipated. Although it can perhaps serve the current practical purposes of exchanging information and carrying on the conversation, it does not always lead to sufficient understanding. Instead, it can sometimes cause further misunderstandings and may not lead to anywhere close to sufficient understanding. Misunderstanding is not always noticed. Usually, the interlocutors just

carry on their interaction without being aware of it. Misunderstanding is more observable in task-solving interactions in Study 2 than in first encounters in Study 1.

Non-understanding is another form of insufficient understanding or understanding problem. Non-understanding occurs when the information presented is not understood at all, for reasons such as lack of access to the information itself or some background knowledge of relevance. Non-understanding cannot serve the current practical purposes of sharing and making sense of the information presented. It does not correspond to what is intended and anticipated. Normally non-understanding is manifested in the speech act of questioning. It is always detected and most often coped with by the interlocutors, and it usually leads to sufficient understanding in the end. It suggests that with consciously experienced, detected, and corrected understanding problems, people often learn very much.

In the next chapter, I will continue to discuss understanding and its problems in this new empirical data with respect to how understanding problems are detected, handled, and resolved in and through interaction and how understanding problems construct sense-making and understanding sharing.



# 9

## Results of analysis of coping with understanding problems

After having reconceptualised understanding and identified various understanding problems, the thesis will find out how people cope with these understanding problems in the empirical data studied. In Chapter 9, one main research question will be investigated. How are understanding problems detected, handled, and resolved in and through interaction? The complexity of understanding in sense-making and meaning co-construction will be investigated in the empirical data from Study 2.

### 9.1 Detection of understanding problems

People are usually good at noticing when a conversation has gone awry as a result of no understanding or misunderstanding (Hirst et al., 1994).

People are in general quite successful in their use of language. That is because they have strategies for coping with their linguistic limitations. If they cannot understand what is being said to them, they seek clarification and try to work things out (Hirst et al., 1994, p. 214).

Since almost all forms of understanding problems are analytically detectable, what is meant here by noticing and detection is only the participant's noticing and detection (see Chapter 4). According to what has been found in the data, there are detected and undetected understanding problems in interactions (see Table 9.1).

Table 9.1. Participants' detected and undetected understanding problems in Study 2.

Participants' detection	Mis U	Partial U	Non U	Total
Detected	10	68	63	141
Undetected	11	1	0	12
Total	21	69	63	153

Table 9.1 shows that all the partial understanding and non-understanding problems are detected except for one instance of partial understanding. Misunderstanding is almost half detected and half undetected.

### 9.1.1 Always detected non-understanding and almost always detected partial understanding

In the data from Study 2, understanding problems in the form of partial understanding and non-understanding are almost always detected by the participants.

#### The only single occurrence of undetected partial understanding

As shown in Excerpt 7, participants Ann and Bengt summarised their answers to the question of socio-cultural differences between China and Sweden. They found that Chinese people use chopsticks but Swedes use knives and forks and also that Swedish people like to eat sweet stuff but the Chinese do not. In Line 4, Bengt uttered that *we can call it cutlery*. In Line 5, Ann did not say anything but only responded with a laugh. It seems that Ann did not sufficiently but only somewhat partially understand what Bengt was talking about, which is likely because she did not know the English word "cutlery". However, perhaps because Ann could almost guess the meaning of the word "cutlery" without knowing exactly the word, Ann did not explicitly mention this understanding problem to Bengt. Ann did not try to seek more information as what she normally does when she encounters partial understanding or non-understanding problems.

#### **Excerpt 7: The only partial understanding instance that is not detected by the relevant participant: extracted from D1d(150121).v, page 8**

- 1 Ann: chopstick
- 2 Bengt: chopsticks yeah
- 3 Ann: yeah (in combination with head nods)
- 4 Bengt: em: ah: we can call it cutlery
- 5 Ann: < laugh > (with embarrassed emotional and attitudinal reaction)
- 6 Bengt: cutlery and then there is also the sweet stuff right
- 7 Ann: yeah (in combination with head nod)

This instance is analytically identified as a partial understanding, because Ann could not write down the point “cutlery” in her notes as she had been doing during the project and the interaction did not proceed as anticipated. For instance, in Line 5, one might expect some kind of signal from Ann about her perception, understanding, and acceptance of Bengt’s suggestion of “cutlery” as one solution to the task. A laughter from Ann is probably not expected in this context. Laughter in communication can be more complicated than one thinks in certain cultural contexts. It is very common that Chinese speakers laugh, chuckle, or smile in an interaction when they feel embarrassed or hesitant (Lu, 2014). From an analytical perspective, Ann’s laughter here (Line 5) expresses her emotional and attitudinal reaction of embarrassment to the English word “cutlery” in Bengt’s earlier utterance (Line 4). This partial understanding was not exhibited clearly by Ann as usual. Bengt is not knowledgeable or aware of various meanings of Chinese laughter, and perhaps Ann is normally a happy person, therefore this partial understanding was not detected by Bengt as usual. This is the only partial understanding case that was not detected by the participants in the data in Study 2.

#### All the other cases of detected partial understanding and non-understanding

Apart from this single case of partial understanding, all the other cases of partial understanding and non-understanding are always detected in the data. This is probably because as long as it is sensed that the interaction contribution (i.e., speech utterance and gestural behaviour) deviates from the intention and anticipation, the communicators usually spend time and effort to reform the interaction in a way that coincides with what is intended and anticipated. It perhaps also has a lot to do with the participants’ friendly and kind attitude towards one another, which is nevertheless in fact the most prominent nature and premise of ordinary communication activities. Partial understanding and non-understanding are usually brought up by one participant and coped with by both participants jointly.

#### **Excerpt 8: Three instances of detected understanding problems: extracted from D2s(150121).f, pages 1–2**

- 1 Ann: something is similar like em do you think swedish people always looks like shame
- 2 Bengt: head forward (non-understanding which shows that Bengt did not get the message
- 3 and he wanted to get more information)
- 4 Ann: e: not shame e: like the how to say in english ah feel ah
- 5 Bengt: ashamed (partial understanding, eliciting for confirmation, shows uncertainty)
- 6 Ann: yeah ashamed it’s a em: (hand finger pointing CPUE/A confirmation like “bingo”)
- 7 Bengt: of what (partial understanding, accompanied with gestural behaviours of eyes
- 8 squeezed and head tilt, which shows disagreement/ question, eliciting for further clarification)
- 9 Ann: character eh people characters like eh always if we are not familiar with each other

10 Bengt: head nod  
 11 Ann: we don't  
 12 Bengt: [talk]  
 13 Ann: [usually] yeah talk with  
 14 Bengt: em  
 15 Ann: [so how to say]  
 16 Bengt: [that is eh:] how to say I'm not sure then then not that open I guess  
 17 Ann: < yeah > maybe  
 18 Bengt: if you sit in a tram  
 19 Ann: ehm  
 20 Bengt: this is people are sitting if this is like two seats two seats two seats  
 21 Ann: yeah [they]  
 22 Bengt: [people would sit] one here [one here one here one here one here one here]  
 23 Ann: [yeah yeah yeah yeah]  
 24 Bengt: and nobody will talk [even]  
 25 Ann: [yeah]  
 26 Bengt: when [standing in line] they won't talk  
 27 Ann: [yeah yeah] [I think]

Excerpt 8 presents three understanding problems that occurred again between Bengt and Ann. At first, Ann said *something is similar like em do you think swedish people always looks like shame* (see Line 1). Bengt did not understand, so he tried to move physically closer to Ann by leaning his head forward which is a typical micro-feedback gestural behaviour expressing uncertainty and eagerness in order to get more information (see earlier in Study 1). This eliciting response from Bengt showed Ann that she needed to provide more information and clear up this non-understanding problem, which was perceived and understood sufficiently by Ann.

Then, Ann started to repair the interaction by saying *e not shame e like the how to say in english ah feel ah* (Line 4). Although Bengt still did not sufficiently understand it, he showed his willingness to cooperate in the meaning repair by suggesting an English word *ashamed* in a rising tone (prosodic feature) to check if it was what Ann meant. This is a case of partial understanding. Bengt's prosodic communicative behaviour not only expressed his uncertainty in sense-making, but also his eagerness to seek further clarification and confirmation so as to promote his partial understanding to a sufficient understanding. Then, Ann confirmed this by saying *yeah ashamed it's a em:* (Line 6) with a complementary gesture of fingers pointing at Bengt, which signals "bingo, that's it".

After this, Bengt became even more confused and wanted more explanations from Ann. Bengt expressed his question and almost disagreement through his vocal-verbal micro-feedback *of what* in Line 7 and its accompanying micro-feedback gestures head tilt and eye squeezed (see Line 8). In particular, Bengt's gestural behaviour showed that he did not accept Ann's opinion or at least felt doubtful about it and

questioned it. Bengt did not sufficiently understand Ann, which also is a partial understanding case. Then, Ann started to repair the interaction by providing more information *character eh people character like eh always if we are not familiar with each other* (Line 9). With the subsequent utterances and examples of how Swedish people sit on the tram and how they communicate in daily life, Bengt showed that he sufficiently understood that what Ann meant earlier was that “Swedish people are very shy”. It seems that all these three understanding problems were caused by the different concepts of this particular English word “shame”. What Ann meant from the start was “shy”, but she misused “shame” instead. Nevertheless, both the first non-understanding problem and the following two partial understanding problems presented in Excerpt 8 have been detected and worked through by the participants in their interaction.

As the empirical data show, both partial understanding and non-understanding are always detected, except for one partial understanding; they are always brought up in the form of a question or suggestion seeking further clarification, explanation, and confirmation; they are always attended to immediately after detection. Non-understanding is expressed by vocal-verbal and gestural behaviours primarily signalling *what did you say, please pardon, sorry, I do not understand, can you explain*, and the like. Partial understanding is exhibited by means of suggestions or questions about what has been partially understood and what needs to be further clarified such as *do you mean this..., is it this...*, and the like.

### 9.1.2 Not always detected: misunderstanding

Misunderstanding, however, is quite different from partial understanding and non-understanding in that it is sometimes detectable and detected but not always.

#### Detected and corrected misunderstanding

When misunderstanding is detected and corrected, detection usually takes place at the same time as misunderstanding has occurred (see Excerpt 9 below). The correction can take short or long time. Also, the detected misunderstanding problems are almost always repaired and promoted to sufficient understanding.

#### **Excerpt 9: Detected and corrected misunderstanding example: extracted from D6d(150204).f, pages 5–6**

- 1 Filippa: head nods (showing agreement) otherwise talk about this one like if the speeches like [emo
- 2 tional or not emotional]
- 3 Enn: [where where where]
- 4 Filippa: on the next page I think it is
- 5 Enn: there (asking for confirmation) additional page (misunderstanding of “the next page” as “ad
- 6 ditional page” in the English language, caused by different concepts of the same English word)

- 7 Filippa: on the next page yeah (correcting)  
 8 Enn: okay yeah (showing realisation) I didn't come here yet  
 9 Filippa: yeah  
 10 Enn: specific [versus natur]  
 11 Filippa: [this one] like if you talk more like emotional  
 12 Enn: m lack of self-controls  
 13 Filippa: like if you like to talk with emotions like  
 14 Enn: yeah

Excerpt 9 above shows that Enn misunderstood Filippa's word *next page* as *additional page* (see Lines 5 to 6) and Filippa detected and corrected it immediately in Line 7 by saying *on the next page yeah*. Then, Enn sufficiently understood it in the way as Filippa intended and anticipated. This misunderstanding problem was likely caused by different concepts of the same English phrase "the next page", and it was detected and corrected immediately after it had occurred, and sufficient understanding was achieved in the end.

### One case of undetected misunderstanding

In the empirical data studied, it has been found to be very common that the participants carried on their interaction without being aware of any misunderstanding problem. Misunderstanding was not detected by the participants (though analytically detected by the researcher).

#### **Excerpt 10: Undetected misunderstanding example: extracted from D11d(150220m).f, pages 11–13**

- 1 Kang: this individual group and hierarchical  
 2 Louise: eh (signalling hesitation)  
 3 Kang: hierarchical  
 4 Louise: ah yes hier hierarchical things  
 5 Kang: m (with head nod)  
 6 Louise: what do you mean in a way (eliciting, showing uncertainty)  
 7 Kang: I mean the similar to the group orientation but it means it's a working it's a working ah culture  
 8 right  
 9 Louise: m  
 10 Kang: then in the in the organisation I think in china you should listen to your [boss]  
 11 Louise: [boss] yeah (with head nods, showing agreement)  
 12 Kang: but here it's like we could have a kind of different thinking  
 13 Louise: yeah (shows agreement)  
 14 Kang: you know we could discuss [it]  
 15 Louise: [it's] more group oriented [you mean]  
 16 Kang: [yeah exactly] (with head nods, showing agreement)  
 17 Louise: eh:  
 18 Kang: no groups is like family  
 19 Louise: m but you discuss it as a team this is how I understand it  
 20 Kang: yes  
 21 Louise: that the you discuss as a team and then you come to a decision as a team rather than eh  
 22 Kang: m (with head nods)

- 23 Louise: your boss tells you what to do  
 24 Kang: yes (with head nods, showing agreement)  
 25 Louise: I have no experience of how it is to be individualistic [because I never worked by myself] like  
 26 that  
 27 Kang: [I know] no (with head shake)  
 28 Louise: I always work in teams so (by Louise "individualism" is regarded as equal to "work alone  
 29 and work independently" and "collectivism" as equal to "work in team", this is a misunderstanding  
 30 possibly caused by different conceptualisations of the notions "individualism" and "collectivism")  
 31 Kang: yes but does it I I mean I can see that here is like more eh chinese way the group no like more  
 32 eh yes chinese way you listen to a boss  
 33 Louise: ehm (with head nods)  
 34 Kang: nut I don't know what about swedish like in the company [how it would work]  
 35 Louise: [but] I think like for example it depends on in what area you are working at cause if you are  
 36 working at a factory and your boss tells you to do [something]  
 37 Kang: [then] you need to it yes [exactly] (showing agreement)  
 38 Louise: [yeah] whereas if you are working like a I don't know (with head nods, signalling  
 39 confirmation)  
 40 Kang: if you are working like a in exactly a group so (...)

In Excerpt 10, participants Kang and Louise were discussing the cultural differences between Sweden and China. Kang came up with *this individual group and hierarchical* in Line 1. Louise did not sufficiently understand it and asked *what do you mean in a way* for more clarification in Line 6. Then, in the following discourse exchanges, Kang and Louise tried to negotiate the meaning of "individualism" and "collectivism". However, in a global context, it is easy to see that a misunderstanding of these two notions has occurred between the participants.

Kang believes that China has more of a collective and hierarchical culture and Sweden has the opposite; whereas, Louise believes that Sweden is a collective country and hierarchical in some areas, because people often work in teams and sometimes need to listen to the boss. Apparently, this misunderstanding is caused by different definitions and conceptualisations of "individualism" and "collectivism" by the participants. This misunderstanding did not seem to bother Kang and Louise in any sense at all. They just carried on their interaction without being aware of it. One thing worth mentioning here is that how Kang's and Louise's conceptualisations of "individualism" and "collectivism" deviated from the reading material provided by the project is not an interactional understanding question and is thus not discussed in greater depth<sup>54</sup> in this thesis.

---

<sup>54</sup> In the reading material provided by the project, it says that *individualism* and *collectivism* refer to social frameworks in which individuals prioritise individual or group needs. In individualistic societies, the primary responsibility of an individual is to take care of him or herself, and his/her individual abilities and characteristics are the primary consideration; whereas, in collective societies, the responsibility to family and groups is most important, the individuals are expected to look out for one another, the ability to fit into groups is more important than individual ability (Connerley & Pedersen, 2005, pp. 41–42).

## One undetected misunderstanding that causes further undetected misunderstanding(s)

In Study 2, it is also found that sometimes one single misunderstanding can result in several more misunderstandings and none of them is detected by the participants during the interaction. This is in line with the findings of misunderstanding in Study 1 (see Section 5.3.2).

### **Excerpt 11: Example of undetected misunderstanding causes further understandings: extracted from D8d(150209).v, pages 8–10**

- 1 Henrik: i think actually the biggest one is the natural versus affective  
2 Guo: yeah yeah that's the most obvious I would say  
...  
3 Henrik: yeah it will be  
4 Guo: so but we have four but shall we eliminate one of them  
5 Henrik: let's destroy one  
6 Guo: yes  
7 Henrik: yeah which one is less the least  
8 Guo: m: (signals thinking)  
9 Henrik: different that's a hard question < chuckle > (signalling uncertainty)  
10 Guo: yes yeah it is  
11 Henrik: i say may be specific versus diffuse or may be individualism versus communitarian (prosody)  
12 Guo: yes i think so (misunderstanding which causes further misunderstanding)  
13 Henrik: yeah (misunderstand.caused by the prior, analytically identified & proved in follow-up interviews)  
14 Guo: that one we should do we should delete it (staring at computer eliciting confirma.) (misunderst. as above)  
15 Henrik: because yeah yeah I think we should skip that one (misu as above)  
16 Guo: yap so  
17 Henrik: alright  
18 Guo: good problem solved  
19 Henrik: high five < chuckle > (hands up to gesture "high five" over the computer)  
20 Guo: high five < chuckle > (hands up to gesture "high five" over the computer) good shall we shall  
21 we knock at the door to notify them

Excerpt 11 shows that Guo and Henrik had almost completed the project task when they realised that they had worked out more solutions than needed so they decided to delete one. In Line 11, Henrik gave a suggestion *i say may be specific versus diffuse or may be individualism versus communitarian* with a slightly rising tone, seeking consensus. Guo mistook this question as a statement and responded *yes i think so* in Line 12. Thus, the first misunderstanding occurred and was undetected by the participants.

Then, surprisingly, in the next utterance (Line 13) Henrik said *yeah* and confirmed. This is a further undetected misunderstanding caused by the prior undetected misunderstanding, and it is analytically identified in accordance with the individual's follow-up interview. Here, evidently, Guo believed that they had agreed on deleting "individualism versus communitarian", whereas Henrik thought they had agreed on taking away "specific versus diffuse".

Following that, Guo said in Line 14 *that one we should do we should delete it* with a gesture of staring at Henrik over the computer, which signals seeking more confirmation, however, without specifying clearly which is his *that one*. This is a continued misunderstanding.

Interestingly, in the next utterance (Line 15), Henrik confirmed Guo's earlier utterance by saying *because yeah yeah I think we should skip that one*, again without clarifying which one is his *that one*. Thus, another misunderstanding caused by earlier misunderstanding(s) occurred. Both Henrik and Guo thought that they had understood each other and they believed that they were talking about the same *that one*. However, it turns out that Guo's *that one* was *individualism versus communitarian* whereas Henrik's *that one* was *specific versus diffuse*.

This is a typical case of one undetected misunderstanding causing further undetected misunderstandings. The first misunderstanding in utterance 12 caused misunderstanding in utterance 13 and then resulted in misunderstandings in utterances 14 and 15; furthermore, none of them was detected by the interlocutors during the interaction.

### 9.1.3 Comparison between partial understanding, non-understanding, and misunderstanding in terms of detection

As regards similarities, the detection of partial understanding, non-understanding, and misunderstanding usually takes place simultaneously or immediately after the understanding problem has occurred. Once an understanding problem has been detected, the interlocutors will start repairing and correcting it jointly. The repair or correction of any understanding problem can take a short or long time and varies between individual situations. The aim is to solve the understanding problem and achieve sufficient understanding. Almost always, these detected partial understanding, non-understanding, and misunderstanding problems are promoted to sufficient understanding in the end.

Regarding differences, partial understanding and non-understanding are almost always detected, whereas misunderstanding is not always detectable and very often undetected. Partial understanding and non-understanding are usually revealed or expressed in the speech acts of question and suggestion; whereas misunderstanding often occurs in statements, which makes it almost blend into the majority of statements that communicate sufficient understanding. These similarities and differences discussed above suggest that misunderstanding plays a more special role than partial understanding and non-understanding in building up sense-making and information sharing.

How understanding problems are handled and resolved will be presented in more detail in the following two sections.

## 9.2 Handling understanding problems

Hirst et al. (1994) have already claimed that the interlocutors usually try to figure out the exact nature of the understanding problem and give additional clarifications, explanations or specifications that can (help) achieve the information sharing.

Participants in a discourse sometimes fail to understand one another, but, when aware of the problem, collaborate upon or negotiate the meaning of a problematic utterance. (Hirst et al., 1994, p. 213)

The empirical data show that the participants handle understanding problems through understanding negotiation and meaning repair, which can be self-initiated or other-initiated<sup>55</sup> (see Table 9.2).

*Table 9.2.* Participants' self-initiated and other-initiated meaning repair of understanding problems in Study 2.

<b>Participants' repair</b>	<b>Mis U</b>	<b>Partial U</b>	<b>Non U</b>	<b>Total</b>
Self-initiated repair	3	4	5	12
Other-initiated repair	7	64	58	129
Total	10	68	63	141

As presented in Table 9.2, most of the understanding problems are handled by other-initiated repair. Only less than a tenth of partial understanding and non-understanding as well as a quarter of misunderstanding instances are handled by self-initiated repair.

---

<sup>55</sup> Definitions of self-initiated and other-initiated meaning repair have been presented in Section 2.9. To summarise, self-initiated meaning repair refers to the repair initiated by the speaking person of the information that causes understanding problems or difficulties; whereas other-initiated meaning repair refers to the repair initiated by the listening person, who has not sufficiently understood the information communicated and who wants further clarification.

### 9.2.1 Self-initiated meaning repair: usually in a form of a statement repeating, modifying, or adding information

As presented earlier in Excerpt 4 (see Section 8.3.2), participant Chun brought up a new topic of “uncertainty avoidance” and Daniel could not find it in the reading material, so Chun specified *the fourth page*. However, it was misperceived and misunderstood as *first page* and *firth* by Daniel. When Daniel’s first misunderstanding of *first page* occurred, Chun immediately detected it and carried out one self-initiated repair by saying *fourth* once more. However, Daniel still did not get it right and mentioned the other misunderstanding *firth*. Again, Chun detected it directly and once more initiated meaning repair by adding information *one two three the fourth paragraph uncertainty avoiding*. Thus, Daniel understood sufficiently and responded with *em: oh*. The understanding problems were immediately detected, self-initiatively repaired and simultaneously corrected by Chun in the interaction sequence.

Similarly, Excerpt 9 (see Section 9.1.2) presents a case where Filippa brought up a topic of “emotion and emotion control” and Enn wanted to know where it was in the reading material, thus asking *where where where*. Filippa responded *on the next page I think it is*, which was misunderstood by Enn as *additional page*. Then, Filippa detected this misunderstanding problem and carried out a self-initiated meaning repair by repeating *on the next page* until they together worked out a sufficiently shared understanding.

In the empirical data studied, it is very common that the participant self-initiates a meaning repair as soon as s/he detects that there is an understanding problem. Self-initiated repair usually takes place in the form of statement, repeating, modifying, or adding information. The repair process can take up more or less a few rounds of discourse exchange. Usually, after a number of utterances and contributions in the interaction, the detected understanding problem can be resolved in a way that is good enough to achieve sufficient understanding.

### 9.2.2 Other-initiated meaning repair: usually in the form of question seeking confirmation, clarification, explanation, or specification

As discussed earlier in Section 8.3.2, Excerpt 6 shows a case of other-initiated meaning repair. The Chinese participant Ann noted that *the swedish wine is so heavy*. Because the Chinese and the Swedish have different concepts of the English word “wine”, the Swedish participant did not understand it sufficiently. So, the other speaker Bengt initiated a meaning repair by repeating the perceived information *okay wine* in the form of a question in order to seek confirmation and further clarification. Ann was brought into this other-initiated meaning repair process after

Bengt's initiation. Later on, after a few rounds of negotiation, the partial understanding problem is resolved and Bengt and Ann have reached a point good enough for further communication.

Also, in Excerpt 8 (see Section 9.1.1), Ann was talking about her opinion of a characteristic of Swedes *em do you think swedish people always looks like shame*. Bengt did not understand it at all, so he initiated a meaning repair by gestural micro-feedback head forward to seek more clarification. This is a case of other-initiated repair. Then, Ann explained that *e: not shame e: like the how to say in english ah feel ah* which was still not sufficiently understood by Bengt. Therefore, he carried out another other-initiated repair by suggesting the English word "ashamed". However, Ann followed up with *yeah ashamed it's a em:* and did not come up with a constructive resolution to this partial understanding problem. Bengt then carried out another other-initiated repair by asking *ashamed of what* together with eliciting gestural micro-feedback of squeezed eyes and tilt head. Because Ann could not find the right English word, she tried to explain what she meant by giving a few examples. Eventually, Bengt understood sufficiently that what Ann was talking about was actually the Swedish people's shy personality. This is a case where the main listener, Bengt, initiated meaning repairs three times and eventually achieved sufficient understanding.

Other-initiated repair is also found very common in the empirical data studied. It is initiated by the interlocutor who is not the initial speaker of the problematic information, and it usually occurs in a form of question in order to seek more confirmation, clarification, explanation, or specification. Like self-initiated repair, other-initiated repair can also take up a few rounds of discourse exchange and in the end lead to sufficient understanding.

### 9.2.3 Comparing self-initiated and other-initiated meaning repair

The majority of partial understanding and non-understanding problems are handled by other-initiated repair, and a very small portion is handled by self-initiated repair.

Concerning the similarities between self-initiated and other-initiated meaning repair, they both take place when one interlocutor detects an understanding problem and s/he is eager to repair and solve it. Also, both self-initiated and other-initiated meaning repair can take up several rounds of discourse exchange and meaning negotiation. The common goal is to resolve the understanding problem and achieve sufficient understanding. In practice, the initial speaker of the information presented that is being repaired is the repair maker (repairer) for both self-initiated and other-initiated repair.

As regards the differences, self-initiated repair usually occurs in the form of a statement, often with the aim of repeating, modifying, or adding information,

whereas other-initiated repair usually occurs in the form of question with the aim of eliciting further confirmation, clarification, explanation, or specification, through for instance, repetition, paraphrase, or reformulation. Also, in self-initiated repair, the initiator of the repair is the initial speaker of the information presented that is being repaired, who is also the leader of the repair and the main repair maker, while in other-initiated repair, the initiator of the repair is the other interlocutor, not the initial speaker of the problematic information, who is primarily listening to the initial speaker's presented information that is being repaired. The initiator of other-initiated repair is the leader of the repair but not the main repair maker. Instead, the initial speaker of the presented information that is being repaired is the main repair maker.

### 9.3 Resolving understanding problems

In this section, whether understanding problems are resolved and whether they construct sense-making and information sharing will be looked at in detail.

*Table 9.3.* Understanding problems in relation to the corresponding sufficient understanding in Study 2. (Note: if the understanding problem is resolved and promoted to sufficient understanding, it is categorised as *resolved*, otherwise *unresolved*.)

Relate to sufficient U	Mis U	Partial U	Non U	Total
resolved	9	68	57	134
unresolved	12	1	6	19
Total	21	69	63	153

Table 9.3 illustrates that in Study 2, the majority of partial understanding and non-understanding problems are resolved by means of meaning repair and some kind of sufficient understanding is eventually achieved. Regarding misunderstanding, however, even with meaning negotiation and repair, there are more instances that are not resolved or promoted to sufficient understanding than those that are resolved and promoted to sufficient understanding.

### 9.3.1 Almost always resolved and promoted to sufficient understanding: partial understanding

In the data studied, partial understanding, apart from the one case presented earlier (see Excerpt 7 in Section 9.1.1), is found to be always resolved and promoted to sufficient understanding through questions and suggestions by means of repetition and paraphrase.

As discussed earlier, when partial understanding occurs, the interlocutor (primarily the listener) has usually partially understood the information communicated. This is primarily due to reasons such as new to knowledge and experience, different cultural or conceptual perspectives of the same concept, and misperception and mispronunciation of a certain word or sentence. The interlocutor (the listener) usually carries out the other-initiated meaning repair to negotiate the meaning of the shared information that is problematic for sufficient understanding. The meaning repair and negotiation is usually performed by means of active follow-up questions and suggestions in the form of repetition or paraphrase of the partially understood information, such as *do you mean this...*, *is it this...*, and the like. The intention is to elicit further confirmation, clarification, explanation, or specification from the initial speaker of the information presented so that the understanding problem can be resolved and sufficiently shared understanding can be promoted. Although a partial understanding problem is initially brought up by the interlocutor who has the understanding problem, it is taken care of collaboratively by both interlocutors in the interaction. Interlocutors interactively construct the sense-making and information sharing. Almost always, a sufficient sharing is achieved in the end.

In Excerpt 12, for example, participants Guo and Henrik were discussing “universalism” and “particularism”, in their own words “rules and equality” and “relationship and inequality”. From Lines 1 to 14, we can see that Guo believed that in China, particularism and relationships were more important than rules and that in Sweden it was the opposite, whereas Henrik believed that in Sweden it was not uncommon that people evaluated and acted out relationships over rules and that there was also corruption and inequality. Then, in Line 15, Guo accepted that it is a kind of universally shared human nature that personal relationships play a role in life and sometimes it can even outweigh societal rules. However, Guo still believed there was a difference between Sweden and China. In Lines 16 and 18, Guo gave examples such as criminals in China may be able to walk away if they know the right persons and Chinese media never reports how the taxes are paid by government leaders, which is very different compared with Sweden.

**Excerpt 12: Partial understanding example: extracted from D8d(150209).v, pages 4–7 (on two occurrences)**

- 1 Guo: i have one about this this universe universalism and particularism I think china is much more depend  
2 on the particularism I mean our laws I mean a lot of things are designed based on depend on particular  
3 persons  
4 Henrik: okay (with head up-nod)  
5 Guo: instead of we have unique judgement around all kinds of people I mean here the law system and the  
6 society treat people much more equally  
7 Henrik: you think sweden is more universalist (using paraphrase, confirming whether information has been  
8 understood correctly and sufficiently)  
9 Guo: yes  
...  
10 Henrik: i don't really agree (with eyebrow frown, signalling disagreement/thinking)  
11 Guo: is that so (with smile and chuckle, asking for more clarification in a friendly way)  
...  
12 Henrik: depending on if you want to get somewhere because I mean there is always small small senses of  
13 corruption everywhere of course the rules are important but if you know the right people then you can do  
14 more or less whatever  
...  
15 Guo: [but focus on] yeah I mean this this is just natural human natures but if you count about like the laws  
16 and that stuff I mean if it's commercial or it's not serious if you kill people  
17 Henrik: yeah  
18 Guo: or if you do some serious crime and if you get contact with people can still walk around [in many  
19 senses]  
20 Henrik: [in] in china (eliciting)  
21 Guo: yeah  
22 Henrik: m (with head nods)  
23 Guo: but this is never possible in sweden I think  
24 Henrik: yeah not this possible yeah yeah true (with head nods, signalling agreement)  
25 Guo: yeah yeah and our medias never covers the personal life of our leaders but it's very possible that the  
26 swedish leaders not paying their fee properly or chuckle  
27 Henrik: ah yeah true true [yeah]  
...  
28 Guo: let me let me give you an example like suppose you want to go or your kids want to go to the best  
29 school in town what's your behaviour  
30 Henrik: if i want to go to the best school in town (not certain, so repeats to ask for confirmation)  
31 Guo: yes what's your action  
32 Henrik: < in sweden > (eliciting)  
33 Guo: yes  
34 Henrik: you get good grades and then pay the fees  
35 Guo: chuckle yeah (showing amusement) but if you are not reaching the line  
36 Henrik: then there is really not so many options  
37 Guo: yeah yeah we have a lot of options so chuckle  
...  
38 Henrik: yeah (signalling agreement) that's true alright let's write it down

Then, Guo introduced another example in Lines 28 to 29 *let me let me give you an example like suppose you want to go or your kids want to go to the best school in town what's your behaviour*. Following that, Henrik paraphrased in Line 30 *if i want to go to the best school in town* with the intention of confirming the correctness of his perceived and understood information. Guo then confirmed this in Line 31 by saying *yes* and asked further *what's your action*. However, Henrik still did not think he had

sufficiently understood Guo's question, so he asked further in the next utterance *in sweden* (Line 32) in order to elicit more complete information. Accordingly, in Line 33 Guo provided another confirmation *yes*. Hence, the partial understanding problem that Henrik detected and initiated himself for further meaning repair and negotiation was jointly solved by Henrik and Guo. In the end, the participants achieved sufficiently shared understanding as anticipated.

### 9.3.2 Often resolved and promoted to sufficient understanding: non-understanding

As presented earlier, when the interlocutor (primarily the listener) has not perceived or understood the information communicated at all, non-understanding occurs. This is very likely because of non-perception and lack of knowledge or experience. The interlocutors of non-understanding problems are always very active and eager to repair the problems. For instance, they elicit more information by asking *sorry, what, what did you say, pardon, or can you say that again*, which are usually related to non-perception. They elicit further clarification by asking *what do you mean, sorry, I didn't understand, can you explain, what's that, or what does that mean*, often because of a lack of knowledge or experience. Often, non-understanding is resolved and promoted to sufficient understanding by means of questions like these.

#### Resolved cases of non-understanding

The one interlocutor, usually the initial speaker of the information presented that is problematic for understanding, repeats, paraphrases, reformulates, or elaborates the information until the other interlocutor has sufficiently understood. Through this joint meaning repair process, the interlocutors often resolve the non-understanding problem and construct a shared understanding collaboratively.

#### **Excerpt 13: Non-understanding example: extracted from D10s(150218).f, pages 2–3 (on two occurrences)**

- 1 Inn: yeah and when we come to university it becomes a very high level that we need to study
- 2 Jenny: head nod
- 3 Inn: a lot
- 4 Jenny: em is there how are the school hours there (eliciting)
- 5 Inn: what do you mean
- 6 Jenny: eh can you for example decide whether you go to school or is it obligatory (eliciting)
- 7 Inn: eh say that again sorry (uncertainty/embarrassment from the prosody, embarrassment from the chuckling)
- 8 the chuckling)
- 9 Jenny: is it do you have to go to the lessons or can you skip some time to time (eliciting)
- 10 Inn: no it's just in the university that you can skip lesson
- 11 Jenny: head nod

For example, Excerpt 13 shows a sequence of interaction between participants Inn and Jenny about the cultural similarities between Sweden and China. Inn said that in China, when it comes to university level, students have to study a lot and very hard (see Lines 1 and 3). Jenny showed her sufficient understanding by means of a gestural micro-feedback of head nod (see Line 2) and further asked *how are the school hours there* in Line 4. However, Inn did not understand Jenny's question at all, so Inn carried out an other-initiated meaning repair in her response *what do you mean* in Line 5. Jenny then elaborated her earlier question as *eh can you for example decide whether you go to school or is it obligatory* (see Line 6). However, Inn still did not understand and responded with another other-initiated repair question *eh say that again sorry* in Lines 7 and 8. Inn's non-understanding was perhaps caused by a lack of perception or not knowing the meaning of *school hours* and *obligatory*. Then, Jenny made another effort to reformulate her question and asked again *is it do you have to go to the lessons or can you skip some time to time* (see Line 9). After that, Inn seemed to have sufficiently understood Jenny's question and responded in her next utterance *no it's just in the university that you can skip lesson* (Line 10). Jenny was satisfied and nodded her head in Line 11. Thus, the non-understanding problems were resolved, and a sufficient understanding was achieved.

**Excerpt 14: Non-understanding example: extracted from D10s(150218).f, pages 3–4 (on two occurrences)**

- 1 Jenny: with the individualism and collectivism
- 2 Inn: which one
- 3 Jenny: on page 42
- 4 Inn: again (with an eliciting prosody)
- 5 Jenny: in the individual societies individuals are expected to take care of themselves eh in collective
- 6 societies individuals are expected to look out for one another
- 7 Inn: m yeah that's what most people say
- 8 Jenny: but how much do you have to take care of yourself in china for example if you lose your
- 9 job do you have something to fall back of
- 10 Inn: yeah we have that (with head nod to show confirmation)
- 11 Jenny: m
- 12 Inn: so I think that's collectivistic societies or (eliciting)
- 13 Jenny: m I think both are (with head nods, signalling agreement)

Similarly, in Excerpt 14, the participants Jenny and Inn were looking for cultural similarities between Sweden and China. Jenny started talking about “individualism and collectivism” (see Line 1). However, Inn did not understand it as Jenny had anticipated, perhaps because of a lack of perception or a lack of knowledge or experience of the particular notions or terminologies. Therefore, Inn initiated an other-initiated meaning repair by asking *which one* to seek more information (see Line 2). Jenny then responded *on page 42* in Line 3. However, Inn still did not understand so

she asked once more by uttering *again* (Line 4) to acquire more clarification and promote her understanding. Then, Jenny explained what *individualism and collectivism*” meant from her point of view (see Lines 5 to 6): *in the individual societies individuals are expected to take care of themselves and in collective societies individuals are expected to look out for one another*. After this, Inn had sufficiently understood what Jenny was talking about and responded that she thought China was more of a collective society (see Line 12). This was later agreed on by Jenny in her next utterance (Line 13) *m I think both are*, that is, Jenny believed that both China and Sweden were collectivistic societies. Inn’s non-understanding problem had now been cleared up and a sufficiently shared understanding was successfully achieved between the participants.

### Unresolved cases of non-understanding

Besides the non-understanding cases that are resolved and promoted to sufficient understanding, it is found that there are 6 out of 63 cases of non-understanding unresolved and not promoted to sufficient understanding (see Table 9.4).

In each of these unresolved cases, the interlocutor (the primary listener) has explicitly expressed the non-understanding problem of the presented information, however, the other interlocutor (the primary speaker) either wanted to ignore it or wanted to fix it with meaning repair but failed (and then gave up and changed the topic). Therefore, sufficient understanding was not achieved.

Table 9.4. Non-understanding cases that are not resolved or promoted to sufficient understanding. (Cases 4–6 occur in sequence in the dialogue.)

Non-U	Dial. No.	Adjacent utterances	Why unresolved
1 Which one is the most obvious one	D13s (150220e).v	Niklas: [is this the same in china with time] Min: <u>[which one is the most obvious one] ah (eliciting)</u> Niklas: is it the same in china with time Min: yeah	Ignored and changed topic
2 Sorry what was the first	D4s (150123).v	Chun: ... so maybe it's another eh similarity Daniel: < sorry what was the first> <u>(with head forward)</u> Chun: < eh I think eh I don't know > maybe skip this one and find another things	Ignored and changed topic
3 Activity vs what	D11d (150220m).f	Louise: <m> and then there were two things we didn't really agree upon which was eh activity orientation and the individual[[ism versus collectivism] Kang: <u>activity vs what</u> Louise: orientation Kang: orientation Louise: and on page 42 individualism and collectivism Chun: yes	Attended, tried repair work, but not repaired, gave up and changed a topic
4 Family is quite different- what did you say	D9d (150218).v	Inn: <1chuckle>1 <2yeah>2 m let me think m m I think family is quite different Jenny: <u>what did you say</u> Inn: <e> in china if we compare it with sweden e e how should I explain	Attended, tried repair work, but not repaired
5 Smile	D9d (150218).v	Jenny: <u>smile (friendliness/patience/uncertainty) (waiting for Cf4 to find words and give more specification)</u> Inn: <m> because family in china we often yeah I have hard to say	Attended, tried repair work, but not repaired
6 Smile	D9d (150218).v	Jenny: <u>smile (same as above smile)</u> Inn: men what do you think about family in oh I think the school in china and in sweden Jenny: head nod Inn: is quite different Jenny: m (with head nod)	Attended, tried repair work, but not repaired, gave up and changed a topic

As seen in Case 4 in Table 9.4, the Chinese participant Inn suggested a topic *I think family is quite different*; however, the Swedish participant Jenny did not understand it and responded with an other-initiated repair *what did you say* to get more clarification. Then, Inn responded *in china if we compare it with sweden e e how should I*

*explain*, which shows that Inn was aware of Jenny's understanding problem and wanted to clarify it but had difficulty giving a clarification and explanation.

After this, in Case 5 in Table 9.4, Jenny did not say anything, merely responding with a smile, which signals that Jenny did not get her non-understanding problem clarified, she wanted more information, and she waited patiently and friendly for Inn to find the right words and organise her thoughts. Inn then responded in the next utterance that *because family in china we often yeah I have hard to say*, which did not resolve Jenny's non-understanding problem but expressed again Inn's difficulty in making further clarification of her own point.

In Case 6 in Table 9.4, Jenny smiled once again without saying anything, which signals her patience, friendliness and also eagerness to get more of an explanation. However, Inn said next *men what do you think about family in oh I think the school in china and in sweden (is quite different)*, which shows that Inn was still aware of the unresolved non-understanding problem and could not fix it herself; instead, Inn wanted to get help from Jenny to repair the problem. However, perhaps because Inn realised that it was not proper to ask Jenny to do it, she changed a topic from "family" to "school" simultaneously in the same utterance.

All these non-understanding problem cases were detected and attended to varying extents, however, the problems were not resolved and sufficient understanding was not achieved. As shown in the data, these unresolved understanding problems did not seem to have an effect on the subsequent interaction. The participants usually chose to move on and switch to a new topic and continued their interaction. It is perhaps because the participants were more driven to solve tasks and that they were eager to complete the tasks within the recommended time.

### 9.3.3 Not always resolved or promoted to sufficient understanding: special role of misunderstanding

Why does misunderstanding play a more special role than the other two forms of understanding problems in conversation? As presented in Table 9.5, more than half the misunderstanding problems in Study 2 are not detected or corrected by the interlocutors. The interlocutors just carry on interacting with one another without being aware of the misunderstanding problems that have occurred. Sometimes, one misunderstanding may even cause further misunderstandings without any of them being noticed.

Table 9.5. Misunderstanding in Study 2 in terms of detection and correction.

Detection and correction	Misunderstanding
Detected and corrected	9
Detected but uncorrected	1
Undetected	11
Total	21

In the empirical data, the misunderstanding problems detected turned out to be resolved or not to promote sufficient understanding, depending on whether the misunderstanding is corrected or not.

In the following subsections, the detected and corrected misunderstanding, the detected but uncorrected misunderstanding, and the undetected misunderstanding will be discussed separately. Then, I will make an attempt to generalise the patterns and features of misunderstanding in the interaction studied.

#### Detected and corrected: always constructive for sufficient understanding

Although the detected misunderstanding may temporarily affect the interaction flow and reveal an understanding problem, it almost always plays a constructive role in building up a sufficiently shared understanding. In particular, the detected and corrected misunderstanding problems are found to be always resolved and promoted to sufficient understanding in the data studied.

It is also found that a once misunderstanding problem is detected by the interlocutors, it is always (except one case, see Excerpt 15 in Section 9.3.3) corrected through meaning repairs and negotiations (see Excerpt 4 in Section 8.3.2 and Excerpt 9 in Section 9.1.2 earlier).

#### Detected but uncorrected: unconstructive for sufficient understanding

As theoretically assumed by Hirst et al. (1994), sometimes misunderstanding is detected but not pointed out or pointed out but not corrected. This type of detected but uncorrected misunderstanding has been found in the empirical data, though with only one occurrence.

#### **Excerpt 15: The only misunderstanding instance that is detected but not corrected: extracted from D1d(150121).v, pages 3–4**

- 1 Ann: I I think the swedish people like / eat eating the sweet food like maybe the candies too sweet
- 2 [too much sugar]
- 3 Bengt: [eat like candy] (Partial U, with gestural micro-feedback of head forward and eyes squeezed)

- 4 Ann: eh sugar  
 5 Bengt: sugar (Partial U, with a gestural micro-feedback of head nod)  
 6 Ann: yeah  
 7 Bengt: we like to eat sugar (Partial U)  
 8 Ann: yeah (with chuckle, Mis U, caused by different concepts of the same English word  
 9 "sugar": Ann thinks sugar equals to sweet stuff, candy, cake; while Bengt thinks sugar is raw sugar)  
 10 Bengt: like just raw sugar (Partial U)  
 11 Ann: eh no just like something | something including the sugar too much sugar (with head shakes)  
 12 Bengt: oh in the tea or coffee or (Mis U, because of different concepts of "sugar")  
 13 Ann: eh like the dessert | and (with gestural micro-feedback gaze up searching for words)  
 14 Bengt: the what (Non-U, with head forward and eyes squeezed)  
 15 Ann: dessert dessert dessert (with head tilt and smile, showing uncertainty and embarrassment)  
 16 Bengt: tidder (misunderstanding, caused by mispronunciation of the English word "dessert")  
 17 Ann: cakes  
 18 Bengt: cake okay (with head nod)  
 19 Ann: yeah  
 20 Bengt: cakes okay alright ...

Excerpt 15 illustrates the only case of detected but uncorrected misunderstanding in the data. Participants Ann and Bengt were discussing the cultural differences between Sweden and China. The Chinese participant Ann suggested a topic of sweet food (see Line 1), but she made her argument by using the English words "candy" and "sugar" (see Lines 1, 2, 4, and 11) which confused the Swedish participant Bengt, who had either a partial understanding problem or a misunderstanding problem (see Lines 3, 5, 7, 10, and 12).

Approaching the end of their meaning repair process, Ann said in Line 13 *eh like the dessert*, but Bengt did not understand at all and wanted to seek more clarification and asked *the what* in Line 14. Ann then uttered *dessert dessert dessert* in Line 15 to explain. However, possibly because of Ann's mispronunciation of the English word "dessert", Bengt misunderstood it as *tidder* and he tried to ask for confirmation in Line 16. The, Ann apparently gave up repairing, and changed the topic from "dessert" to *cakes* (see Line 17). Then, Bengt showed his sufficient understanding of "cakes" in Lines 18 and 20, although it is very likely Bengt had still not sufficiently understood "dessert". Although this detected but uncorrected misunderstanding case is constructive for the global conversation, it is regarded as unconstructive for the local sufficient understanding in this thesis.

Although normally the interlocutors are friendly to each other and usually have the joint goal of accomplishing successful communication, the detected but uncorrected misunderstanding, which is unconstructive for sufficient understanding, still occurs in interactions.

### Undetected: deconstructive to promote sufficient understanding

There are also many undetected misunderstanding cases in the data studied. The participants continued communicating with one another without any problem throughout the interaction. As seen from Excerpt 10 and Excerpt 11 presented in Section 9.1.2, the interlocutors misunderstood one another although they erroneously believed that they had sufficiently understood each other and that they had already achieved a sufficiently shared understanding. Such cases were also revealed in the follow-up interviews.

From the perspective of interaction construction, anything that occurs in an interaction is constructing the interaction. Misunderstanding is as constructive for interaction as sufficient understanding or any other form of insufficient understanding in that all forms of understanding build up the context and the interaction discourse. However, incorrect sense-making and information sharing that is unintended and unanticipated are not constructive but deconstructive for promoting shared understanding. Only the sense-making and information sharing that is more or less in harmony with what is intended and anticipated is constructive in promoting sufficient understanding.

With misunderstanding that is undetected and thus uncorrected, the interaction may still continue and be completed without the interlocutors being disturbed; however, the interlocutors are not able to achieve a sufficiently shared understanding in particular as regards this misunderstood information and its related subsequent information communicated. In this sense, undetected and uncorrected misunderstanding is somewhat deconstructive when building up shared understanding. Therefore, it makes more sense to investigate whether there is any pattern or linguistic sign for misunderstanding (see more discussions in Section 9.5.4).

## **9.4 Summary of the detection, handling, and resolving of understanding problems**

To summarise, in Study 2, all the non-understanding problems and all the partial understanding problems except one are detected, and more than half (11 out of 21 cases) of the misunderstanding problems are undetected (see Table 9.6). The only undetected partial understanding was analytically observed as a result of the participant's unexpected laughter at the English word "cutlery", which was likely due to the participant not knowing the word "cutlery" (see Excerpt 7 in Section 9.1.1).

Table 9.6. Overview of detection and repair of understanding problems in Study 2.

Detection	To Suff. U	Meaning repair	Mis U	Partial U	Non U	Total
Detected	Resolved	Self-initiated and repaired	3	4	5	12
		Other-initiated and repaired	6	64	52	122
	Unresolved	Other-initiated but unrepaired	1	0	6	7
Undetected		No repair	11	1	0	12
Total			21	69	63	153

As presented in Table 9.6, among the detected understanding problems, all the detected partial understandings and all the detected misunderstandings except one are repaired and promoted to sufficient understanding in the end. Only around 10% (6 out of 63 cases) of non-understandings are not repaired or promoted to sufficient understanding (see excerpt details in Section 9.3.2).

It is found that the majority of them are fixed by other-initiated repairs and a small portion of the repaired understanding problems are the results of self-initiated repairs (122 versus 12 cases, see Table 9.6). Furthermore, there is an even smaller portion of detected understanding problems that are coped with by other-initiated repairs to a varying extent but remain eventually unrepaired (7 occurrences in total including 1 misunderstanding and 6 non-understandings). It turns out that all the self-initiated repairs succeeded in meaning repair and the majority of other-initiated repairs achieved their goals too.

Only one misunderstanding with other-initiated repair failed in meaning repair. This occurred because of the mispronunciation of “dessert” by one participant. This instance of misunderstanding was detected by the other participant, who consequently initiated with a meaning repair. Although this meaning repair was performed by the first participant, it was not successfully repaired as anticipated. In the end, the participants gave up repairing and changed the topic (see Excerpt 15 in Section 9.3.3). Equally important, 6 non-understanding problems were tackled with other-initiated repairs, and they are not repaired successfully either (see excerpt details in Section 9.3.2). The reasons why they are not repaired as intended are more or less because the initial speaker of the information presented, which is problematic for understanding, deliberately wants to ignore or accidentally ignores it, or wants to fix it but fails after making an effort and then decides to give up (see details presented in Table 9.6). Usually, these unrepaired understanding problems end up with a change to a new topic and are forgotten and “disappear” in the discourse. They may construct the conversation, however, they are not constructive when it comes to

achieving sufficient understanding of the misunderstood or non-understood information.

## 9.5 Discussion

Study 2 shows that different forms of understanding problems occur somewhat differently when it comes to their detection, handling, and resolving. It is easier for the interlocutors to detect partial understanding and non-understanding problems than misunderstandings. This is consistent with earlier studies (e.g., Weigand, 1999; Lu, 2014).

In this section, the similarities and differences between these three identified and analysed understanding problems will be discussed in terms of how they are detected, handled, and resolved. Why misunderstanding plays a more special role than partial understanding and non-understanding in interaction will be discussed.

### 9.5.1 Micro-feedback, contextual account, and relevance indicator

As discussed earlier, partial understanding and non-understanding are usually revealed through questions and micro-feedback expressions such as *do you mean this...*, *is it this...*, *sorry*, *what*, *what did you say*, *please pardon*, *can you say that again*, *what do you mean*, *what's that*, *can you explain*, and the like. However, misunderstanding is usually not exposed through micro-feedback expressions such as these but by some relevant communicative actions and responsive actions (see also Lu, 2014). Interpretation and identification of misunderstanding has to be built upon contextual and relevance accounts. This is in line with earlier theories of understanding in conversation (e.g., Gumperz, 1982; Schegloff, 1987; Tannen, 1993; Bauman & Briggs, 1990; Couper-Kuhlen, 2001).

### 9.5.2 Construction to interaction and sense-making

It has been found that partial understanding and non-understanding construct both interaction and sense-making. They usually achieve a sufficiently shared understanding after meaning repairs. Although misunderstanding is another form of insufficient understanding and it constructs interaction discourse just like partial understanding and non-understanding, it does not construct sense-making or information sharing as intended and anticipated and it does not lead to sufficiently shared understanding when it is uncorrected or undetected. Misunderstanding makes sense of the shared information, but it does not make sense as intended and anticipated. The sense and meaning made in misunderstanding deviates from the intended sense and meaning. This is along the lines of studies by, for example, Zaefferer (1977), Schegloff

(1987), Weigand (1999), Verdonik (2010), and Lu (2014). Misunderstanding can be constructive when promoting sufficient understanding only if it is detected and corrected during the interaction and the interlocutors are friendly and kind towards each other.

### 9.5.3 Social communicative activity and communication task

Apart from that partial understanding was not found in Study 1 of first encounters but only emerged in Study 2 of task-solving collaboration, misunderstanding was also found to be neither detected nor corrected in Study 1 but only detected, repaired, and resolved in Study 2. The detected and resolved misunderstanding cases account for nearly half of all the occurrences of misunderstanding. This may be because in Study 2 the participants are more task-solving driven than Study 1, and that in order to solve the given learning assignments, the participants must engage in intensive high quality exchanges of information which make high demands on mutual understanding and knowledge (re)constructing. Accomplishing complex learning tasks has made achieving sufficient understanding more substantial and thus has made their interaction more likely revealing of insufficient understanding. This is why friendly and kind interlocutors may have been able to detect and resolve more understanding problems in Study 2 than Study 1. This is in line with the theories of Wittgenstein's (1958) language games, Austin's (1962) and Searle's (1969) speech acts, Grice's (1975) meaning and implicature, Levinson's (1979) activity type, Allwood's (2013) activity-based communication analysis, Gumperz' (1982) relevant aspects of activity type, and Linell's (2010) communicative activity type. Studying language use and communication understanding is dependent on the social communicative activity.

### 9.5.4 Is there any linguistic "sign" for misunderstanding?

Based on the empirical data studied, misunderstanding has been found to be very likely caused by reasons such as misperception of an English word (e.g., Excerpt 4 in Section 8.3.2), different concepts of the same English word (e.g., Excerpt 3 in Section 8.3.2 and Excerpt 9 in Section 9.1.2), different perspectives of same concept (e.g., Excerpt 10 in Section 9.1.2), misperception of an English word or sentence (e.g., Excerpt 11 in Section 9.1.2), and mispronunciation of an English word (e.g., Excerpt 15 in Section 9.3.3). Among them, linguistic and conceptual differences (i.e., different concepts of the same English word or different perspectives of the same concept) are the main reasons for misunderstanding occurrences (see Section 8.3.4).

In these cases, the misunderstood information is usually repeated (e.g., Excerpt 4 in Section 8.3.2 and Excerpt 15 in Section 9.3.3), paraphrased (e.g., Excerpt 9 in Section 9.1.2), or responded to with unanticipated actions (e.g., Excerpt 3 in Section 8.3.2) in its next utterance. Therefore, in addition to closely checking the local and even global discourse context of relevance, the communicators should be aware that a problem of misunderstanding might have occurred when exchanged information is repeated, paraphrased, or responded to with unanticipated actions. These linguistic patterns might be useful for the design of communication technology applications such as systems for understanding recognition and automatic detection of misunderstanding.

## 9.6 Conclusion of Chapter 9

This chapter mainly addresses the question of how understanding problems are detected, handled, and resolved in and through interaction. How understanding problems construct sense-making and information sharing is investigated.

The analysis shows that understanding problems are detected when the interlocutor realises that the on-going interaction deviates from what is anticipated and intended. Normally, the interlocutors are friendly and kind to each other and loyal to their communication activity. They are engaged in accomplishing their communication task, and they are eager to clear up understanding problems and promote shared understanding. Although usually understanding problems are initiatively detected and brought up by one interlocutor, they are always coped with by interlocutors interactively by means of either self-initiated or other-initiated meaning repairs.

Once an understanding problem is detected, meaning repair and negotiation are usually initiated immediately and may continue in a large number of discourse exchanges until the detected understanding problem is resolved or given up by the participant for further repair. The detected and repaired understanding problems are found to be constructive for both interaction and sufficient understanding. In such cases, the meaning repair and negotiation are aligned with the intended and anticipated sense-making and information sharing. On the other hand, the undetected or detected but unrepaired understanding problems are not constructive for promoting sufficient understanding, although they are still constructive for the conversation.

Partial understanding is almost always detected and usually handled by means of other-initiated repair questions that suggest, repeat, or paraphrase what has been partially understood and what needs to be further confirmed or clarified. Partial understanding is usually revealed by micro-feedback expressions such as *do you mean*

*this...*, *is it this...*, and the like. The detected partial understanding problems are always repaired and thus promoted to sufficient understanding.

Non-understanding is usually handled by means of other-initiated repair questions. It is commonly expressed through vocal-verbal and gestural micro-feedback that signals *what did you say, please pardon, sorry, I do not understand, can you explain*, and the like with the aim of eliciting further clarification and explanation. Non-understanding problems are always detected and most often, but not always, repaired and promoted to sufficient understanding. Only in a few cases, non-understanding is not promoted to sufficient understanding, because the interlocutor either deliberately ignores it or makes repair without success and then gives up.

Misunderstanding is commonly not detected by the interlocutors. The detected misunderstanding is handled more often by other-initiated meaning repairs than self-initiated ones. Detected misunderstanding is almost always repaired and corrected and thus promoted to sufficient understanding of what has been intended and anticipated. Misunderstanding is found to play a more special role than partial understanding and non-understanding in constructing sense-making and information sharing. It is primarily because misunderstanding is very often undetected and the undetected misunderstanding is constructive for interaction but usually unconstructive and sometimes deconstructive for the intended sense-making and information sharing. In addition to that a local or a global context can help to identify misunderstanding, the analysis also suggests that when the presented information is repeated, paraphrased, or responded to with unanticipated actions, a misunderstanding has probably occurred.

In the next chapter, comparisons will be made of the occurrence, detection, handling, and resolving of understanding and understanding problems in the two contexts of face-to-face and video-mediated communications.

# 10

## Results of comparing understanding between FTF and VMC

In Chapter 9, understanding and understanding problems have been studied in terms of their detection, handling, and resolving in the communication data focused on. In this chapter, understanding and its problems will be investigated with a focus on comparing understanding between face-to-face and video-mediated communication (henceforth FTF and VMC). The research question of what similarities and differences there are between face-to-face and video-mediated communication in the occurrence, detection, handling, and resolving of understanding and understanding problems will be explored.

### 10.1 Overview of the data

Table 10.1 presents the occurrences of different understandings in these two different communication media with main analytical focuses on micro-feedback and meaning repair. The size of the data corpus is presented in time, words, and utterances.

Table 10.1. Overview of the data in Study 2<sup>56</sup>. (Note: Word = number of words; Utt. = number of utterances; Suff. U = sufficient understanding; Mis U = misunderstanding; Partial U = partial understanding; Non U = non-understanding; Decimals are rounded to 0.01.)

	Dialogue number	Length			Understanding problems				
		Time	Word	Utt.	Suff. U	Mis U	Partial U	Non U	Total
<b>VMC</b>	D1d(150121).v	06:27	752	121	87	3	6	3	12
	D4s(150123).v	08:18	965	133	85	3	4	4	11
	D5s(150204).v	08:41	1327	159	106	1	1	7	9
	D8d(150209).v	09:30	1177	134	101	4	7	1	12
	D9d(150218).v	11:45	842	110	72	0	1	7	8
	D12s(150220m).v	09:50	1479	197	124	1	5	4	10
	D13s(150220e).v	11:20	1459	192	120	0	6	5	11
	D16d(150309).v	05:04	525	100	77	0	2	0	2
	D17d(150310).v	10:39	1059	143	117	1	1	0	2
	D20s(150314).v	15:57	1230	157	137	1	7	2	10
<b>Total of VMC</b>		1:37:31	10815	1446	1026	14	40	33	87
<b>Per minute</b>			110.90	14.83	10.52	0.14	0.41	0.34	0.89
<b>Per utterance</b>		02:31	7.48		0.71	0.01	0.03	0.02	0.06
<b>FTF</b>	D2s(150121).f	09:35	1110	179	99	2	5	2	9
	D3d(150123).f	12:35	1377	226	135	0	2	3	5
	D6d(150204).f	04:28	741	103	69	1	0	1	2
	D7s(150209).f	08:52	1165	186	120	0	2	1	3
	D10s(150218).f	08:33	472	62	41	0	2	2	4
	D11d(150220m).f	12:59	1744	257	142	2	3	11	16
	D14d(150220e).f	07:08	874	148	104	0	4	1	5
	D15s(150309).f	08:42	804	161	119	1	2	3	6
	D18s(150310).f	05:20	539	91	73	1	3	4	8
	D19d(150314).f	13:45	1028	160	123	0	6	2	8
<b>Total of FTF</b>		1:31:57	9854	1573	1025	7	29	30	66
<b>Per minute</b>			107.17	17.11	11.15	0.08	0.32	0.33	0.72
<b>Per utterance</b>		02:18	6.26		0.65	0.00	0.02	0.02	0.04
<b>Total of entire data</b>		3:09:28	20669	3019	2051	21	69	63	153
<b>Per minute</b>			109.09	15.93	10.82	0.11	0.36	0.33	0.81
<b>Per utterance</b>		03:35	6.85		0.68	0.01	0.02	0.02	0.05

<sup>56</sup> Both the communication tasks and the communication situations are counterbalanced in this study (see more in Chapter 4 Method). In Table 10.1, the dialogues are numbered according to the time when they were recorded, and are named as follows: first the dialogue number, then the communication task (d=differences and s=similarities) and the recording date and last the communication media (v=vmc and f=ftf) (see also Section 4.2.6).

As presented in Table 10.1, there are 2051 occurrences of sufficient understanding and 153 occurrences of understanding problems identified in Study 2 through micro-feedback and other related responsive actions, primarily, meaning repair.

## 10.2 Descriptive statistics of the data

Descriptive statistics of the comparisons between VMC and FTF will be presented in the following section.

### 10.2.1 Comparison between VMC and FTF

The distributions of data on the level of dyads were examined in histograms. The distributions of conversation length, number of words, number of utterances, and occurrences of sufficient understanding appear to be similar, as shown in Figures 10.1 to 10.4.

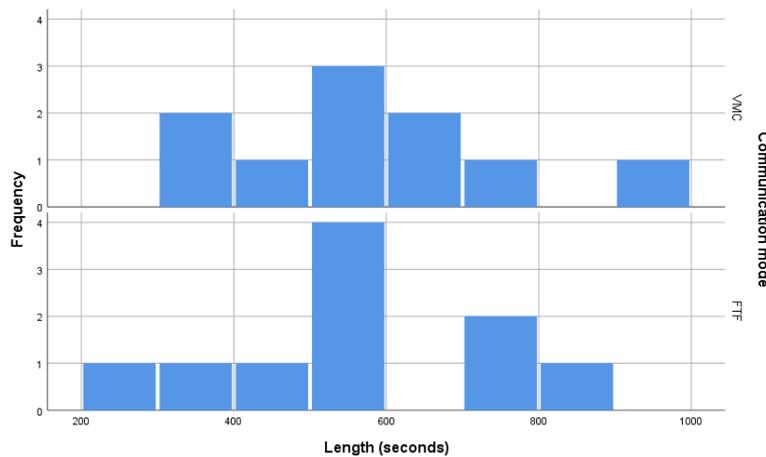


Figure 10.1. Histogram of the length of conversations shown for VMC and FTF.

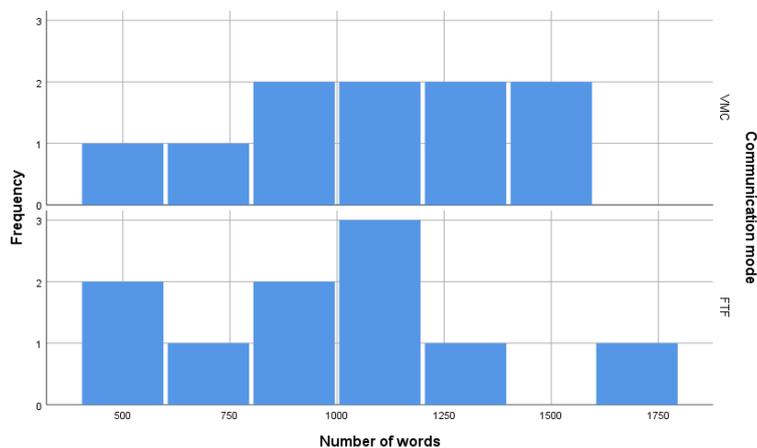


Figure 10.2. Histogram of the number of words of VMC and FTF conversations.

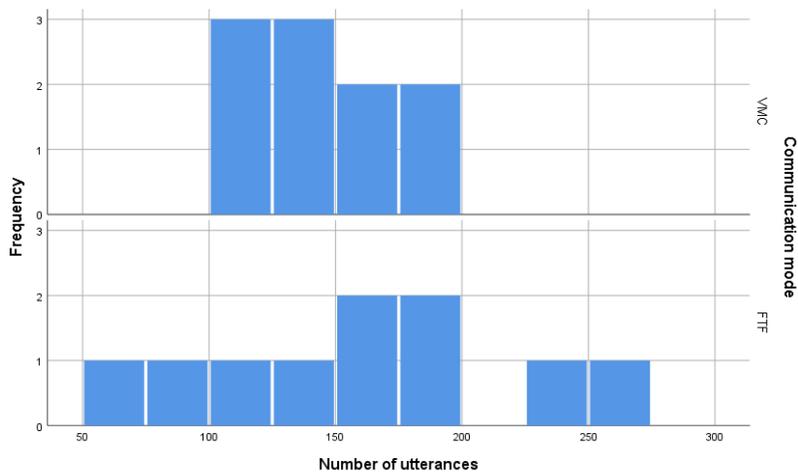


Figure 10.3. Histogram of the number of utterances of the conversations in VMC and FTF.

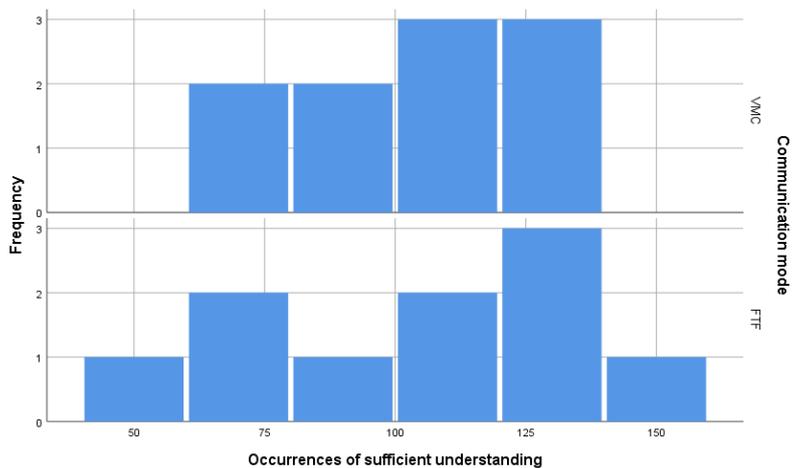


Figure 10.4. Histogram of the occurrences of sufficient understanding in the conversations, shown for VMC and FTF.

Equally important, Figure 10.5 shows that the occurrences of understanding problems could differ slightly between VMC and FTF. However, a later inferential test (Chi-square) did not show any difference (see Section 10.4). Perhaps this could be because of a low peak in VMC and a high peak in FTF (see the histogram in Figure 10.5) influencing the inferential test. Without these data at the edges of the distributions (two dialogues in VMC and one dialogue in FTF), it would perhaps turn out that VMC has more understanding problems than FTF. However, with the data at hand, drawing such a conclusion is too uncertain.

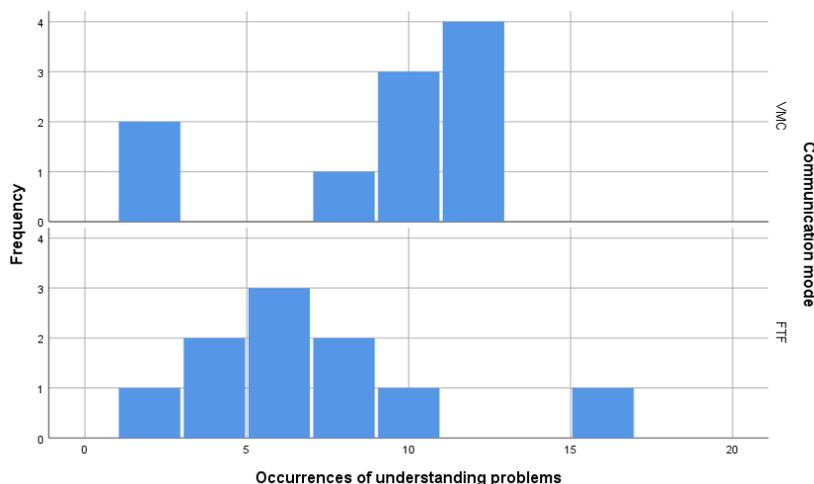


Figure 10.5. Histogram of the occurrences of understanding problems in the conversations, shown for VMC and FTF.

### 10.2.2 Comparison between the collaborative tasks

In this section, the data will be compared with respect to the two different collaborative tasks, namely picking out cultural differences and similarities between Sweden and China (see Appendix G).

Table 10.2 shows that the differences in length (in seconds) go in both directions with half the conversations being shorter and the other half longer. Also, the situation is the same concerning the number of words, the number of utterances, and the occurrences of understanding problems. Half the conversations have more in these frequencies and half have fewer. This is also similar when it comes to the occurrences of sufficient understanding in that about half the conversations have more occurrences and about half have fewer occurrences.

It seems that the data do not show any difference with regard to the length, the numbers of words, and the number of utterances in the conversations, when comparing the collaborative tasks. Either there are no differences in these aspects between the conversations, or around half of the conversations contain more occurrences while around half contain fewer, which evens out the result.

Table 10.2. Differences in length, number of words and number of utterances between the two collaborative tasks (Note: the collaborative tasks of picking out cultural differences and similarities between Sweden and China are named as differences-task and similarities-task in the following tables and figures).

<b>Differences (i.e., freq. of differences-task minus freq. of similarities-task)</b>					
<b>Dyad</b>	<b>Length (s)</b>	<b>Words</b>	<b>Utterances</b>	<b>Suff. U</b>	<b>U problems</b>
1	-188	-358	-58	-12	3
2	257	412	93	50	-6
3	-253	-586	-56	-37	0
4	38	12	-52	-19	0
5	192	370	48	31	4
6	189	265	60	18	6
7	-252	-585	-44	-16	0
8	-218	-279	-61	-42	0
9	319	520	52	44	-6
10	-132	-202	3	-14	-2

Table 10.3 presents the occurrences of sufficient understanding and understanding problems in the conversations of differences-task vs. those of similarities-task. Although it is difficult to infer any pattern from Table 10.3, it provides the bases for later scatterplot analysis (see Figure 10.6).

Table 10.3. Occurrences of sufficient understanding and understanding problems in the two tasks.

<b>Dyad</b>	<b>Differences-task</b>		<b>Similarities-task</b>	
	<b>Suff.Und</b>	<b>Und.Prob</b>	<b>Suff.Und</b>	<b>Und.Prob</b>
1	87	12	99	9
2	135	5	85	11
3	69	9	106	9
4	101	3	120	3
5	72	8	41	4
6	142	16	124	10
7	104	11	120	11
8	77	6	119	6
9	117	2	73	8
10	123	8	137	10

Because the occurrences of sufficient understanding and understanding problems do not seem to differ between the two collaborative tasks, further investigations are

made as follows. Whether there is a linear relationship and its strength between the occurrences of understandings (i.e., sufficient understanding and understanding problems) and the collaborative tasks (i.e., differences-task and similarities-task) is measured by Pearson correlation coefficient  $r$ .

Occurrences of sufficient understanding showed a small tendency to be correlated to the tasks (Pearson  $r = .286$ ,  $p = .423$ )<sup>57</sup>, while occurrences of understanding problems showed a stronger tendency to be correlated to the tasks (Pearson  $r = .457$ ,  $p = .184$ ) (although neither were statistically significant).

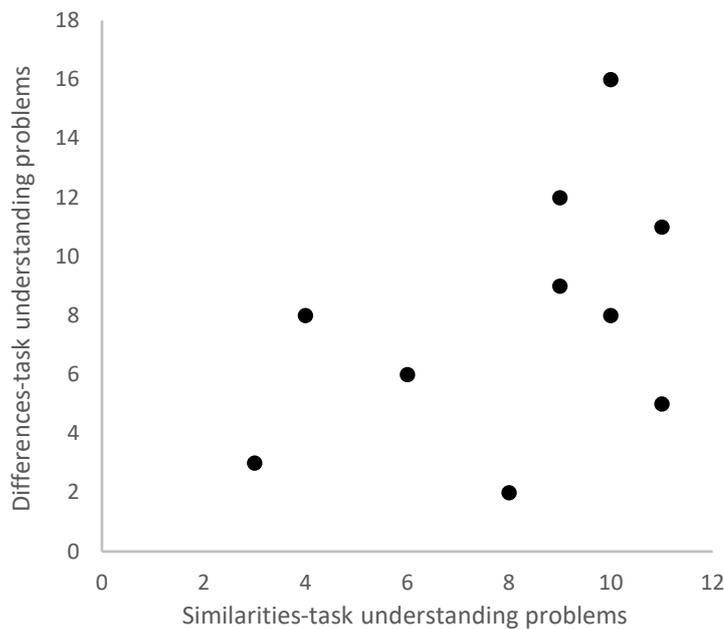


Figure 10.6. Scatterplot of occurrences of understanding problems for the difference-task and the similarities-task, showing the tendency of a linear trend.

Figure 10.6 illustrates the correlations between the occurrences of understanding problems and the collaborative tasks (i.e., difference-task and similarities task, as noted in Table 10.2). Such correlations would mean that the occurrences of sufficient understanding and understanding problems are dependent on the properties of the

---

<sup>57</sup> The value of  $r$  is always between +1 and -1. To interpret its value: 0 to .30 indicates a weak uphill (positive) linear relationship; .30 to .50 indicates a moderate uphill (positive) linear relationship; .50 to .70 indicates a strong uphill (positive) linear relationship; .70 to 1.0 indicates a perfect uphill (positive) linear relationship.

dyads, and not caused by the tasks themselves. Further statistical and qualitative analyses will be presented in the following sections.

### 10.3 Basic difference between VMC and FTF in length

The overview of the data in Table 10.1 (presented in Section 10.1) shows that the difference in length between VMC and FTF is around five minutes<sup>58</sup>. The length of the conversations in VMC and FTF was compared using a repeated measures *t* test (see reasons in Section 4.2.10). There was no significant difference in length for VMC ( $M = 585.1$  s,  $SD = 180.99$  s) and FTF ( $M = 551.7$  s,  $SD = 188.7$  s),  $t(9) = 0.468$ ,  $p = .651$ .

Additionally, Table 10.1 shows how many words the participants have communicated in each dialogue of VMC and FTF. The number of words in the conversations in VMC and FTF was compared using a repeated measures *t* test. There was no significant difference in number of words for VMC ( $M = 1081.5$ ,  $SD = 312.99$ ) and FTF ( $M = 985.4$  s,  $SD = 386.77$ ),  $t(9) = 0.748$ ,  $p = .474$ .

Furthermore, as can be seen from Table 10.1, each dialogue consists of different numbers of utterances. The number of utterances of the conversations in VMC and FTF was compared using a repeated measures *t* test. There was no significant difference in number of utterances for VMC ( $M = 144.6$ ,  $SD = 32.23$ ) and FTF ( $M = 157.3$ ,  $SD = 60.08$ ),  $t(9) = -0.69$ ,  $p = .508$ .

These three tests suggest that VMC is as efficient as FTF in terms of word deployment and turn shift.

### 10.4 Occurrences of understandings in relation to communication media

First, comparing VMC with FTF, there are 1026 versus 1025 instances of sufficient understanding and 87 versus 66 cases of understanding problems, respectively (see Table 10.1). So, does FTF have more sufficient understanding than VMC, and does VMC have more understanding problems than FTF? Second, when it comes to misunderstanding, there are almost twice as many cases in VMC compared with FTF, with frequencies of 14 compared to 7 and 0.14 versus 0.08 per minute (see also Table 10.1). As regards partial understanding, there are 40 cases in VMC and 29 instances

---

<sup>58</sup> This 0:00:05 is short compared to the entire data's length of 3:09:28. What can happen during five minutes in a conversation? Based on this empirical data, in general, within five minutes people can communicate roughly 550 words, 100 utterances, and 80 demonstrated cases of understanding and its problems.

in FTF, with a frequency of 0.41 versus 0.32 per minute. Regarding non-understanding, Table 10.1 shows that both VMC and FTF have similar frequencies, with 33 cases and 0.34 per minute in VMC versus 30 cases and 0.33 per minute in FTF. So, is there any association between the occurrences of different forms of understanding problems and the communication media? Does communication media technology affect understanding?

In order to see if frequencies of understanding type (sufficient understanding, misunderstanding, non-understanding, and partial understanding) were associated with communication mode (VMC and FTF), a 4 x 2 Chi-square test of independence was carried out (see reasons in Section 4.2.10). No association was found using raw frequencies ( $\chi^2(3) = 4.011, p = .26$ ) or frequencies per time unit ( $\chi^2(3) = 3.645, p = .302$ ) (frequency per 100 minutes was used in order to yield integers for the statistical calculation). The result shows no difference between VMC and FTF concerning the occurrences of sufficient understanding, misunderstanding, partial understanding, or non-understanding.

To be specific, as regards sufficient understanding, the result suggests that FTF does not have significantly more sufficient understanding than VMC. In VMC, people can achieve sufficient understanding as much as in FTF. The communication technology of VMC does not affect people's sufficient understanding. Furthermore, concerning understanding problems, the result suggests that VMC does not have more understanding problems than FTF. In FTF, people can encounter as many understanding problems as they do in VMC. As regards misunderstanding, the result shows that both VMC and FTF have a similar probability of a misunderstanding occurrence. Regarding partial understanding, it is suggested that people have a similar likelihood of achieving partial understanding in both VMC and FTF. For non-understanding, the result shows that people's non-understanding problems occur equally often in both VMC and FTF. That is, the result of the statistical test shows that VMC does not have significantly more understanding problems of any type than FTF. Communication technology as studied in this thesis does not have a negative impact on communication understanding.

## **10.5 How are understanding problems detected, handled, and resolved in VMC and FTF?**

In the following sections, VMC and FTF will be compared in detail with respect to the detection, handling, and resolving of understandings. Because some frequencies are too small to conduct meaningful statistical tests, the comparison of the detection,

handling, and resolving of understandings between VMC and FTF will be based on the actual raw frequencies and the comparison only applies to this data.

### 10.5.1 Detection of understanding problems in VMC and FTF

Detection of understanding problems is defined and coded from the participant's perspective, and the analysis is carried out by means of an analytical approach (see Chapters 2 and 4). Table 10.4 shows the participants' detected and undetected understanding problems in the data studied. In all, the majority of the understanding problems are detected (92%), and only a small portion are undetected (8%). There are 91% of the understanding problems are detected in VMC and 9% are undetected, whereas 94% of the understanding problems in FTF are detected and 6% are undetected in FTF.

*Table 10.4.* Participants' detected and undetected understanding problems in VMC and FTF. (Mis=misunderstanding; Partial=partial understanding; Non=non-understanding; percentage is rounded to integer.)

Participants' detection	VMC				FTF				Total
	Mis	Partial	Non	Total	Mis	Partial	Non	Total	
Detected	7	39	33	79 (91%)	3	29	30	62 (94%)	141 (92%)
Undetected	7	1	0	8 (9%)	4	0	0	4 (6%)	12 (8%)
Total	14	40	33	87 (100%)	7	29	30	66 (100%)	153 (100%)

Although the total number of misunderstanding cases in VMC is double that in FTF, there is a common tendency to have similar frequencies of detected and undetected misunderstandings within the same communication situation. For instance, VMC has 7 detected and 7 undetected misunderstanding instances, and FTF has 3 detected and 4 undetected misunderstanding instances. Also, Table 10.4 shows that in VMC almost all the partial understanding problems are detected, except in one single case. In FTF, all partial understanding instances are detected. Furthermore, non-understanding problems are all detected in both VMC and FTF.

### 10.5.2 Handling of understanding problems in VMC and FTF

As can be seen from Table 10.5, 129 out of 141 detected understanding problems (91%) are handled with other-initiated meaning repair, and 12 out of 141 (9%) are handled with self-initiated meaning repair in all the data. All of the meaning repairs

are primarily performed by the initial speaker of the information associated with additional understanding problems (see more in Section 2.9).

*Table 10.5.* Self-initiated and other-initiated meaning repair with or without success in handling the understanding problems studied. (Mis=misunderstanding; Partial=partial understanding; Non=non-understanding; percentage is rounded to integers.)

Participants' repair	VMC				FTF				Total
	Mis	Partial	Non	Total	Mis	Partial	Non	Total	
Self-initiated repair and repaired	0	4	1	5 (6%)	3	0	4	7 (11%)	12 (9%)
Other-initiated repair and repaired	6	35	27	68 (86%)	0	29	25	54 (87%)	122 (86%)
Other-initiated repair but unrepaired	1	0	5	6 (8%)	0	0	1	1 (2%)	7 (5%)
Total	7	39	33	79 (100%)	3	29	30	62 (100%)	141 (100%)

Table 10.5 shows that there are almost twice as many cases of detected understanding problems that are repaired by self-initiated meaning repair in FTF than in VMC, with a frequency of 11% versus 6%. Both VMC and FTF have similar numbers of other-initiated repaired understanding problems (86% and 87%, respectively). What is more, VMC has three times more detected understanding problems that are unrepaired although with other-initiated meaning repair efforts than FTF, with a percentage of 8% compared to 2%.

Most of the detected partial understanding and non-understanding cases in both VMC and FTF are repaired through other-initiated meaning repair. Interestingly, all the detected misunderstanding cases in VMC are handled with other-initiated repair, whereas all the detected misunderstanding cases in FTF are handled with self-initiated meaning repair. This may be because that interlocutors are to some extent more self-aware of the occurrences of misunderstandings in FTF than in VMC, which suggests a higher interdependency (with a focus on reactions, see details in Footnote 25 in Section 2.10.1) and interactivity (with focuses on both actions and reactions, see Footnote 14 in Section 2.4) between people in FTF than in VMC.

### 10.5.3 Resolving understanding problems and promoting sufficient understanding

The repaired understanding problems are found to be always resolved and thus eventually promoted to sufficient understanding, whereas the unrepaired understanding problems are unresolved and thus sufficient understanding is not promoted. Table 10.6 illustrates that through meaning repair and negotiation, 88% (134 out of 153) of the understanding problems are promoted to sufficient understanding and 12% are not (19 out of 153).

*Table 10.6.* Understanding problems resolved or unresolved in relation to sufficient understanding in VMC and FTF. (Suff. U = sufficient understanding; Mis = misunderstanding; Partial = partial understanding; Non = non-understanding; percentage is rounded to integer.)

Relate to Suff. U	VMC				FTF				Total
	Mis	Partial	Non	Total	Mis	Partial	Non	Total	
Resolved	6	39	28	73 (84%)	3	29	29	61 (92%)	134 (88%)
Unresolved	8	1	5	14 (16%)	4	0	1	5 (8%)	19 (12%)
Total	14	40	33	87 (100%)	7	29	30	66 (100%)	153 (100%)

In FTF, almost all the understanding problems are promoted to sufficient understanding, with only one non-understanding case and four misunderstanding instances as exceptions, whereas in VMC, there are in total 14 cases, which are not promoted to sufficient understanding, including 8 misunderstanding instances, 1 partial understanding instance, and 5 non-understanding instances.

As presented in Table 10.6, 84% (73 out of 87 cases) of the understanding problems in VMC are resolved and promoted to sufficient understanding and 16% (14 out of 87 cases) of them are unresolved, whereas in FTF, the percentage of repaired understanding problems that are promoted to sufficient understanding is comparatively higher. There are 92% (61 out of 66 cases) understanding problems are resolved and promoted to sufficient understanding and only 8% (5 out of 66) are unresolved in FTF. Obviously, understanding problems in VMC, with a percentage of 16%, are double those in FTF with a percentage of 8%.

### 10.5.4 Why not promoted to sufficient understanding?

As found earlier, the participants have successfully repaired 134 occurrences of the understanding problems and achieved sufficient understanding in the end. Only 19 cases of understanding problems were not repaired and thus not promoted to sufficient understanding. Table 10.7 shows that there are three reasons why these 19 understanding problems are not promoted to sufficient understanding in the data studied.

*Table 10.7.* Reasons why a few understanding problems are not promoted to sufficient understanding. (Mis = misunderstanding; Partial = partial understanding; Non = non-understanding)

Why not promoted	VMC				FTF				Total
	Mis	Partial	Non	Total	Mis	Partial	Non	Total	
Undetected by participants	7	1	0	8	4	0	0	4	12
Ignored, changed topic	0	0	2	2	0	0	0	0	2
Attended with repair, failed	1	0	3	4	0	0	1	1	5
Total	8	1	5	14	4	1	1	5	19

First of all, some understanding problems are not detected by the participants. They include seven misunderstandings and one partial understanding in VMC and four misunderstandings in FTF. They were not detected by the participants and consequently not repaired or promoted to sufficient understanding.

Second, the participants chose to ignore the understanding problems and change the topic. There are only two non-understanding cases of this sort that only occurred in VMC. It can be analytically seen that the participant deliberately did not want to repair the understanding problem; rather, the participant wanted to ignore it and switch to another topic.

Third, the participant tackled the understanding problem with meaning repair but failed to repair it and sometimes the participant decided to give up repairing and switched to a new topic. There are three non-understanding problems and one misunderstanding of this sort in VMC and only one non-understanding case like this in FTF.

## 10.6 Discussion

In the data on the comparison of the occurrence of understandings, no significant difference is found between FTF and VMC interactions, although in some cases the

non-significant results of statistical tests could be due to low frequencies of the data. Nevertheless, the details of how understanding problems are coped with have shed light on some differences between FTF and VMC (with a focus only on the audio and video chat mode, see description in Chapter 4).

### 10.6.1 Better understanding in FTF than in VMC?

The results from Study 2 show that FTF does not have significantly more sufficient understanding than VMC. Although earlier studies have found that VMC's communication channel is relatively narrow and the media richness is comparatively low (Sins et al., 2011) whereas FTF has more sources of information that regulate interaction, perception, and understanding (Straus, 1997; Driskell & Radtke, 2003; Friedman & Currall, 2003; Paul et al., 2004), the data suggest that these discrepancies do not result in FTF interactions having significantly more sufficient understanding than VMC. Communicating in person does not necessarily result in more sufficient understanding than distance communication using video technology. This supports Anderson's (2006) finding that mediating technology has little effect on the interaction. Video communication technology does not reduce the likelihood of sufficient understanding's occurrence.

### 10.6.2 More understanding problems in VMC than in FTF?

The results from Study 2 also show that VMC does not have significantly more understanding problems than FTF. This does not correspond with the results of earlier studies, for example, that interlocutors in VMC have more trouble in understanding others' responses (Kiesler & Sproull, 1992) and are more likely to experience ambiguity and understanding problems (Shin, 2005). Instead, the data suggest that there are not significantly more understanding problems in VMC than in FTF. This is in line with Anderson's (2006) finding that mediating technologies have little effect on achieving understanding in interactions.

It seems that Kiesler and Sproull (1992) and Shin (2005) focus on investigating the linguistic channels and the communicative and interactive behaviours in VMC versus FTF. However, Anderson (2006) and the present analysis focus on investigating not only the linguistic behaviours but also the outcome of team problem-solving cooperation in VMC versus FTF. Although Anderson (2006) and this analysis differ in that Anderson (2006) compares two-party and three-party VMC and FTF interactions in map task-solving and this analysis compares dyadic VMC and FTF interactions in learning question's task-solving, they are similar in that both studies allow clear measurement of who knew, understood, said, and did what and also how and

to what extent the task-problems were solved. Both studies aim to identify video mediating technology's impact on achieving understanding and communication through participants' interdependent task-solving cooperation, including both task-solving procedures and task-solving results. In other words, video communication technology does not increase the likelihood of occurrences of understanding problems in interaction.

### 10.6.3 More difficult to achieve sufficient understanding and communication goal in VMC

The present analysis has also found that 91% of the understanding problems are detected in VMC and 94% in FTF. Further, 84% of the understanding problems in VMC are repaired and promoted to sufficient understanding, whereas in FTF, the rate is 92%. On the other hand, 16% of the understanding problems in VMC and 8% in FTF are unresolved. Furthermore, almost all the unrepaired non-understanding problems occurred in VMC; both only undetected partial understanding and only detected but unrepaired misunderstanding occurred in VMC; there are more undetected misunderstanding cases in VMC than FTF (see excerpts presented in Section 9.3). These suggest that it is more difficult to detect and repair understanding problems, achieve sufficient understanding, and accomplish the communication goals in VMC than in FTF. These results are in line with earlier theories that VMC is more constrained to achieve common ground and understanding than FTF (Olson & Olson, 2000; Clark, 1996; Clark & Brennan, 1991) and that compared to FTF, interlocutors in VMC have more difficulties achieving communication goals (cf. Anderson, 2006) or have to work harder to achieve them (Whittaker, 2003).

### 10.6.4 Detection of misunderstanding

The data show that almost all the undetected understanding problems are misunderstandings (except for the single partial understanding caused by Chinese laughter). Misunderstanding is more difficult to detect than partial understanding and non-understanding. This supports the earlier finding in Study 1 (see also Lu, 2014). This is also one of the reasons why misunderstanding plays a more special role than partial understanding and non-understanding in discourse interaction (see also Section 9.3.3).

Additionally, it is found that there are more cases of misunderstanding undetected in VMC than in FTF, with a frequency of 7 versus 4 out of in total 21 cases of misunderstanding observed in the entire data. Furthermore, all the detected misunderstanding problems are corrected in FTF but not all are corrected in VMC. These

results seem to correspond to the earlier study by Thompson and Coovert (2006) in which they find communication technologies often lead to difficulties in detecting and correcting misunderstandings.

### 10.6.5 Other-initiated vs. self-initiated meaning repair for misunderstanding

Data from Study 2 have shown that 56% of the participants' detected understanding problems are solved through other-initiated meaning repair, and 44% by self-initiated meaning repair. All these other-initiated and self-initiated repairs are self-performed (see Section 2.9). This is consistent with the earlier theory that self-performed repair is more prevalent than other-performed repair in discourse and conversation (Schegloff, Jefferson, & Sacks, 1977).

Study 2 has also found that in both VMC and FTF, most of the partial understanding and non-understanding cases are solved through other-initiated repair. However, all the detected misunderstanding cases in VMC are found having been solved by other-initiated repairs; whereas all the detected misunderstanding cases in FTF are solved by self-initiated repairs. This can possibly be explained by interlocutors having higher interdependency when interacting in FTF than in VMC, which is consistent with the finding of Stone and Posey (2008). Also, the exchange of auditory and visual communication cues such as prosody and gesture are more self-evident in FTF than VMC, which may make it more possible to self-notice and repair an insufficient understanding of the shared information in relation to what is intended by oneself. This turned out to be in line with a similar claim made by Driskell and Radtke (2003). It is more likely that with more FTF communication resources, the initial speaker of the misunderstood information becomes self-aware of the misunderstanding problem fairly quickly and initiates a meaning repair accordingly.

## 10.7 Conclusion of Chapter 10

In this chapter, a comparative study has been conducted to investigate the similarities and differences between face-to-face and video-mediated communication in the occurrence, detection, handling, and resolving of understanding problems.

Regarding the occurrence of understanding problems, the data show that VMC does not have significantly fewer sufficient understanding or more understanding problems than FTF. That is, video communication technology does not have any impact on people's understanding in interactions: it neither decreases sufficient understanding nor increases understanding problems. The data suggest that these is no

association between the different forms of understanding problems and the communication media.

As regards the detection of understanding problems by comparing the frequencies of detected and undetected cases, the data suggest that it is slightly more difficult to detect understanding problems in VMC than in FTF.

Concerning the handling of understanding problems, almost double the number of detected understanding problems are repaired by self-initiated meaning repair in FTF than in VMC. This suggests that when people communicate in person they have higher self-awareness of the interactivity about what is going on, what has been intended and anticipated, and what is deviated from. This self-awareness of the interactivity helps interlocutors handle understanding problems. Most of the detected partial understanding and non-understanding cases in both VMC and FTF are repaired by means of other-initiated meaning repair. This means that partial understanding and non-understanding are more often detected by the other speaker than by the initial speaker of the information presented. Also, VMC and FTF have similar frequencies of other-initiated repaired understanding problems. This is perhaps because interlocutors in VMC are as loyal, friendly, and kind to each other and as eager to fix understanding problems and achieve communication goals as they are in FTF. Further, all the detected misunderstanding cases in VMC are tackled without success or resolved with success by means of other-initiated repairs, whereas all the detected misunderstanding cases in FTF are resolved by self-initiated repairs. This is probably due to the higher interdependency and interactivity in FTF and VMC.

Regarding the resolving of understanding problems, the detected understanding problems including all three types are much more often resolved and promoted to sufficient understanding in FTF than in VMC. Also, even with other-initiated repair efforts, VMC has three times more detected understanding problems that are unrepaired than FTF. Almost all the unrepaired non-understanding problems occurred in VMC rather than FTF. This suggests that understanding problems in VMC are comparatively more difficult to be repaired and promoted to sufficient understanding than those in FTF. There are more cases of misunderstanding problems undetected in VMC than in FTF, which means that misunderstanding is not only difficult to be observed and detected by the interlocutors in interaction, but also more difficult to be observed and detected in video technology-mediated communications than face-to-face situations.

In other words, although FTF does not have advantages over VMC in terms of the occurrences of sufficient understanding and understanding problems, it still provides better chances of detecting, handling, and resolving understanding problems than does VMC.



# 11

## Discussion and summary of Study 2

This chapter presents a discussion and conclusion of Study 2 in this thesis. First, a review of the research purpose and research questions in Study 2 will be presented. Then, the main empirical findings of the analyses in Study 2 will be summarised. Third, contributions and implications of Study 2 will be discussed. Next, critical reflections on the analyses and research limitations of Study 2 will be discussed. Accordingly, suggestions for some possible future studies will be presented.

### 11.1 Review of the research purpose and research questions in Study 2

Study 1 in this thesis, on investigating micro-feedback in relation to understanding in FTF first encounters, was found to be limited in several respects, especially concerning the types of understanding problems and communication contexts. Thus, Study 2, with expanded data of a more complex design of communication activity in the varying communication media of FTF and VMC, was conducted.

Study 2 expanded the research on understanding in real-time communication, and aimed to examine how understanding problems are coped with by acquainted interlocutors in relation to not only micro-feedback but also other responsive interactions, primarily, meaning repair in an educational activity with collaborative learning tasks.

Study 2 comprises three empirical analyses, Chapters 8, 9 and 10, addressing three research questions as shown below (one chapter each):

- RQ4: What are understanding and understanding problems in social communication?
- RQ5: How are understanding problems detected, handled, and resolved in and through interaction?
- RQ6: What similarities and differences are there between face-to-face and video-mediated communication in the occurrence, detection, handling, and resolving of understanding and understanding problems?

Study 2 in this thesis was based on 20 audio- and video-recorded FTF and VMC dyadic intercultural communication dialogues in the English lingua franca between Swedish and Chinese participants who were acquaintances. Their communication task was to solve some educational learning assignments collaboratively. Three empirical analyses were conducted to explore the concept and the phenomenon of understanding and to identify different forms of understandings, with respect to their occurrence, detection, handling, and resolving. The analytical focuses were on micro-feedback and meaning repair, using an interactional approach based on theories of social communicative activity type, meaning and implicature, contextualisation, and relevance. These analyses resulted in a contribution to more nuanced insights into how understanding and understanding problems are coped with in real-time communication.

## **11.2 Summary of the results from Study 2**

The empirical data in Study 2 have shown that sufficient understanding occurs most frequently in interactions, followed by partial understanding and then non-understanding; misunderstanding occurs with the lowest frequency. This section will synthesise the analytical results of Study 2 to answer the research questions.

### **11.2.1 Regarding reconceptualisation of understanding**

What are understanding and understanding problems in social communication? Study 2 (Chapter 8 primarily) has identified a more nuanced classification of understanding in the empirical data—partial understanding. The theoretical analysis in Chapter 8 has provided an insight into understanding: when understanding corresponds to what is intended and anticipated it becomes sufficient understanding and leads to successful communication, whereas when understanding deviates from what is intended and anticipated it becomes insufficient understanding and results in miscommunication. Partial understanding, misunderstanding, and non-understanding

have been identified as insufficient understandings and understanding problems in the data analysed.

Table 11.1 presents the criteria used in this study to classify understanding in interaction, primarily including whether it serves the current practical purposes of information sharing, sense-making, and continuing interaction, in which speech act it is exhibited, how it is related to the interaction intention and anticipation, how the interlocutors feel about it, and how it is related to the intended sufficient understanding.

*Table 11.1.* Reconceptualising understanding and understanding problems.

	<b>Sufficient U</b>	<b>Insufficient U</b>		
		<b>Partial U</b>	<b>Mis U</b>	<b>Non U</b>
Serving the current practical purposes	serves	partially serves	can serve	does not serve
Exhibited in the speech acts of	declaring and persuading statements	eliciting questions, suggest	declaring and persuading statements, similar as sufficient understanding	eliciting questions
In relation to the intention and anticipation	corresponds	partially corresponds	incorrectly related or deviates	does not correspond
Interlocutors feel or experience	content and feel good enough to proceed	want to seek further confirmation, clarification, explanation, or specification, almost always detected and handled	believes that one has understood correctly and even sufficiently, usually not aware of it, undetected	want to seek more clarification, explanation, always detected and coped with
In relation to the intended Suff. U	achieves	developing stage towards achieving sufficient understanding and almost always succeeds	does not always achieve sufficient understanding and can lead to further misunderstandings	usually leads to sufficient understanding

### 11.2.2 On detecting, handling, and resolving understanding problems

How are understanding problems detected, handled, and resolved in and through interaction? The empirical analysis in Chapter 9 explores this question and investigates how understanding problems construct sense-making and information sharing.

Table 11.2 presents the way in which understanding problems are revealed and then detected, whether they are handled by means of other- or self-initiated meaning repairs, whether they are resolved and then promoted to sufficient understanding, if unresolved what the reason is, and its constructive role in interaction and the intended sense-making and information sharing.

Table 11.2. Detecting, handling, and resolving understanding problems.

	Insufficient U		
	Partial U	Mis U	Non U
Revealing	usually revealed by micro-feedback expressions such as <i>do you mean this...</i> , <i>is it this...</i> , and the like	not easily revealed, a local or a global context can help to identify misunderstanding; also when the presented information is repeated, paraphrased, or responded to with unanticipated actions, a misunderstanding may have occurred	usually revealed through vocal-verbal and gestural micro-feedback that signals <i>what did you say</i> , <i>please pardon</i> , <i>sorry</i> , <i>I do not understand</i> , <i>can you explain</i> , and the like
Detecting	almost always detected	not always detected, very often undetected	always detected
Handling	usually handled by other-initiated repair questions which suggest, repeat, or paraphrase the presented information	handled more often by other-initiated meaning repairs than self-initiated ones	usually handled by other-initiated repair questions with the aim of seeking further clarification and explanation
Resolving	detected partial understanding problems are always resolved and promoted to sufficient understanding	detected misunderstanding is almost always resolved and corrected and thus promoted to sufficient understanding	most often but not always resolved and promoted to sufficient understanding
Reason of why detected but unresolved	none	interlocutor makes repair without success and thus gives up	the interlocutor either deliberately ignores it or makes repair without success and thus gives up
Constructive role	constructive for both interaction and the intended sense-making and information sharing	constructive for interaction but usually unconstructive for the intended sense-making and information sharing	constructive for both interaction and the intended sense-making and information sharing

### 11.2.3 Regarding comparing understanding between FTF and VMC

What similarities and differences are there between face-to-face and video-mediated communication in the occurrence, detection, handling, and resolving of understanding and understanding problems? Chapter 10 presents the findings.

Table 11.3 provides a synthesised presentation of the results of Chapter 10 about the comparison of the similarities and differences in the occurrence, detection, handling, and resolving of understanding and understanding problems between VMC and FTF.

*Table 11.3.* Comparing understandings in VMC vs. FTF on the occurrence, detection, handling, and resolving.

	<b>Similarities</b>	<b>Differences</b>
Occurrence	no significant difference in the frequencies of sufficient understanding and understanding problems	slightly higher raw frequencies of sufficient understanding in FTF and fewer understanding problems in VMC
Detection	majority of understanding problems are detected by the interlocutors during the interaction	<ol style="list-style-type: none"> <li>1. slightly more difficult to detect understanding problems, with a lower frequency of the detected and a higher frequency of the undetected respectively in VMC than FTF</li> <li>2. more cases of misunderstanding problems undetected in VMC than FTF</li> <li>3. in the entire data, the only undetected partial understanding occurred in VMC instead of FTF</li> </ol>
Handling	<ol style="list-style-type: none"> <li>1. both are self-repaired by the initial speaker of the information that is associated to the later understanding problems</li> <li>2. partial understanding and non-understanding are often handled by other-initiated meaning repair</li> </ol>	<ol style="list-style-type: none"> <li>1. almost double the number of self-initiated meaning repairs in FTF than those in VMC</li> <li>2. all the detected misunderstandings in VMC are tackled or resolved with other-initiated repair; whereas, all the detected misunderstandings in FTF are resolved with self-initiated meaning repair</li> </ol>
Resolving	majority of understanding problems are resolved by the interlocutors and thus sufficient understanding is achieved in the end	<ol style="list-style-type: none"> <li>1. understanding problems in VMC are more difficult to be resolved and promoted to sufficient understanding than FTF (e.g., VMC has more understanding problems unresolved than FTF has, and almost all the unresolved non-understanding problems occurred in VMC)</li> <li>2. more cases of misunderstanding problems undetected thus also unresolved in VMC than FTF</li> <li>3. in the entire data, the only detected but unresolved misunderstanding occurred in VMC and not in FTF</li> </ol>

#### 11.2.4 Generalisation

Study 2 in this thesis reconceptualises understanding in social interaction and discusses and compares how understanding problems are detected, handled, and resolved in face-to-face and video-mediated interactions by focusing on micro-feedback and meaning repair.

Based on the results from Study 1, Study 2 contributes to a more nuanced classification of understanding, that is, partial understanding. The data show that sufficient understanding occurs most frequently in conversation followed by partial understanding and then non-understanding with misunderstanding occurring with the lowest frequency. In the data, all the detected understanding problems are handled with meaning repair, which is either self- or other-initiated but always self-performed.

Partial understanding is almost always detected and usually handled by means of other-initiated repair in the form of questions. It is usually revealed by micro-feedback expressions such as *do you mean this...*, *is it this...*, and the like. The detected partial understanding problems are always repaired and thus promoted to sufficient understanding.

Non-understanding is usually handled by means of other-initiated repair in the form of questions. It is commonly expressed through vocal-verbal and gestural micro-feedback that signals *what did you say*, *please pardon*, *sorry*, *I do not understand*, *can you explain*, and the like. Non-understanding problems are always detected and most often repaired and promoted to sufficient understanding. They are not always repaired, because the interlocutor deliberately ignores them or makes repairs without success.

Misunderstanding is found to play a more special role than partial understanding and non-understanding when constructing sense-making and information sharing. First, it is very often undetected. Second, although it is constructive for conversation, it is sometimes unconstructive for the intended sense-making and information sharing. A local or a global context can help to identify misunderstanding. Also, misunderstanding often occurs when information is repeated, paraphrased, or responded to with unanticipated actions. The detected misunderstanding is handled more often by other-initiated repair than self-initiated repair, and it is almost always repaired and promoted to sufficient understanding.

Video mediating technology does not seem to affect understanding. The FTF interactions studied do not have more sufficient understandings or fewer understanding problems than the VMC interactions. However, people have higher interdependency and interactivity in FTF than in VMC. It is easier to detect and repair understanding problems in FTF than in VMC. In the empirical data, the only undetected

partial understanding, the only detected but unrepaired misunderstanding, and almost all the unrepaired non-understandings occurred in VMC rather than in FTF. It is difficult to observe misunderstanding in interaction, and it is even more difficult to observe misunderstanding in VMC than in FTF. There are almost double the self-initiated meaning repairs in FTF compared to VMC, and all the detected misunderstandings in FTF are resolved with self-initiated meaning repair whereas in VMC all of them are tackled or resolved with other-initiated repair.

As understanding is an interactive and situated phenomenon shown by the use of language and the construction of discourse, the social communicative activity type (Wittgenstein's (1958) *language games*, Allwood's (1976) *behaviour, situation, and meaning types*, Levinson's (1979) *activity type*, Allwood's (2013) *activity-based communication analysis*, and Linell's (2010) *communicative activity type*), the implication of intended meaning and anticipated reaction (Grice's (1975) *meaning and implicature*), the context of interaction (Gumperz' (1982) *contextualisation*), and the discourse of relevance (Sperber and Wilson's (1986) *relevance theory*) should all be taken into account when studying understanding in communication.

In general, Study 2 in this thesis applies the earlier result of Study 1 about the relation between micro-feedback and understanding in order to analyse understanding and its problems in FTF and VMC. Although micro-feedback plays an important role in the communication exchange process and is one type of evidence of showing acceptance and understanding of the information presented, it does not say everything about sense-making and information sharing in spontaneous communication. Participants also draw other resources, for example, meaning repair, in coping with understanding issues.

The empirical findings are dependent on the communication context. For example, in a context where the interlocutors have conflicting goals, how micro-feedback are related to understanding and what various levels of micro-feedback and understanding are involved or revealed in interaction could possibly be different from what has been described in the present study. Study 2 only proposes a slightly different interactional approach from Study 1 and a more practical way of studying and accounting for understanding in normal social activities such as online and flexible learning, digital communication, business cooperation, video conferencing, and the related virtual agents' animations.

### **11.3 Contributions and implications of Study 2**

Theoretical and practical contributions and implications of Study 2 will be discussed in this section.

### 11.3.1 Smile, chuckle, and laughter in relation to understanding problems

Study 2 has found that partial understanding and non-understanding problems are sometimes expressed through gestural micro-feedback smile and multimodal micro-feedback chuckle and laughter, which primarily signal emotions of friendliness, politeness, uncertainty, and embarrassment. Such partial understanding and non-understanding problems are usually detected and attended to with meaning repairs by the interlocutors, perhaps because of their loyalty and kindness to one another and their joint interactivity. However, these understanding problems are not always repaired or promoted to sufficient understanding, mainly because of reasons such as the initial speaker of the information, which is associated to the understanding problem, does not want to repair or wants to but fails to repair and then gives up repairing.

As found in Study 2, another reason is that these understanding problems, which are related to smile, chuckle, and laughter, are sometimes not detected by the interlocutors and thus not attended to or repaired. First, this is likely because of cultural or individual differences in employing and interpreting smile, chuckle, and laughter in intercultural and interpersonal communication (Lu, 2014). For instance, as identified by Hayakawa (2003), a non-humorous laughter (a person laughs in a situation where there is nothing that could be considered funny) is common in particular in Japanese and Chinese cultures. According to Hayakawa (2003), functions of laughter can include promoting conversation, easing tension in conversation, and covering up problems (e.g., embarrassment and incomprehension). The third function of covering up problems corresponds to the finding in this study, that is, the participants seem to smile, chuckle, or laugh when they have understanding problems. It is not possible to generalise this finding in the Chinese or Swedish culture, considering the size of the data studied in this thesis. However, the analysed data show that the Chinese participants have a tendency to smile, chuckle, and laugh more frequently than the Swedish participants when they do not understand or act as (they believe they are) expected to express emotions like politeness, uncertainty, and embarrassment. Second, micro-feedback expressions of smile, chuckle, and laughter occur very frequently when signalling sufficient understanding (see also Lu, 2014), thus making it more difficult for the interlocutors to distinguish the differences and observe the understanding problems.

### 11.3.2 Cues for detecting understanding problems

In Study 2, partial understanding and non-understanding problems have been found to be almost always revealed through certain vocal-verbal and gestural micro-feedback. For instance, vocal-verbal expressions *do you mean this...*, *is it this...*, and the like usually reveal partial understanding, and vocal-verbal and gestural expressions signalling *what did you say*, *please pardon*, *sorry*, *I do not understand*, *can you explain*, and the like usually exhibit non-understanding.

Additionally, in Study 2 it was found difficult to detect misunderstanding in interaction, which is in line with Lu (2014), and also it was even more difficult to detect misunderstanding in VMC than in FTF interactions. However, Study 2 also found that repetitions and paraphrases of the presented information and unanticipated responses to it can be cues for the occurrences of understanding problems, especially for misunderstandings. Earlier studies of Bertrand and Goujon (2017), Allenmark, Hsu, Roussel, and Waszak (2015), and Weber (1993) have mentioned that repetitions and unanticipated responses are related to mutual understanding and its problems but paraphrases. The present study suggests that when the presented information is paraphrased, repeated, or responded to with unanticipated actions, an understanding problem and very often a misunderstanding may have occurred.

### 11.3.3 Performance, initiation, and outcome of the meaning repairs

Regarding the performance of the meaning repair, results from Study 2 have shown that all the meaning repairs in both FTF and VMC were performed by the initial speaker of the information presented, which is problematic for achieving sufficient understanding. The speaker is the main repairer. This is likely because the speaker of the information presented feels more responsibility than the hearer to correct and repair the understanding problem. This result is in agreement with Schegloff, Jefferson, and Sacks (1977), that self-performed repair is more prevalent than other-performed repair. Other-performed repair, which according to Schegloff, Jefferson, and Sacks (1977) occurs frequently in adult-child interactions, did not occur in the data studied in this thesis.

The initiation of meaning repair can, however, be self-initiated or other-initiated. Study 2 has found that in both FTF and VMC partial understanding, misunderstanding, and non-understanding are often handled by other-initiated meaning repairs rather than self-initiated ones. This is perhaps because normally interlocutors are loyal, friendly, and kind to each other in both FTF and VMC situations, and they are engaged in their joint interactivity. As soon as they reveal the deviation of sense-making from what is intended and anticipated, they want to correct it quickly. This

is also in line with earlier studies by, for example, Schegloff, Jefferson, and Sacks (1977), that interlocutors are adequate self-monitors and self-correctors as a condition of competence for social interactions in real lives.

Study 2 has also found that the outcome of meaning repair can either be successful or not. The detected understanding problems are overwhelmingly frequently repaired (95%), and only a few cases are unrepaired (5%) in the data studied. This is either because the initial speaker of the information that causes problems for understanding does not want to repair it or the speaker wants to but fails repair then abandons any further repairing. Perhaps this has something to do with their communication task of solving some learning assignments together within a given time period. The interlocutors did not want to spend time or effort on repairing the understanding problem, unless they judged it to be important and necessary for them to proceed with the communication and complete the task.

#### 11.3.4 Higher interdependency and interactivity in FTF than in VMC

As found in Study 2 in this thesis, there are almost double the number of self-initiated meaning repairs in FTF (11%) than in VMC (6%). All the detected misunderstandings in VMC are tackled or resolved with other-initiated repairs; whereas all the detected misunderstandings in FTF are resolved with self-initiated repairs. These all suggest that people have higher interdependency and interactivity in FTF than in VMC, which is in agreement with a study by Stone and Posey (2008). Besides, it means that people have higher self-monitoring competence in FTF than in VMC. Because there are more communication resources such as auditory and visual communication cues for prosody and gesture in FTF than in VMC (see Driskell & Radtke, 2003), there is a higher possibility in FTF than in VMC of the speaker self-noticing and repairing an insufficient understanding of his or her presented information in relation to what he or she has intended and anticipated.

Study 2 has also found that compared to FTF, it is more difficult to resolve understanding problems in VMC and promote them to sufficient understanding. There are more unrepaired understanding problems in VMC than in FTF. Almost all the unrepaired non-understanding problems occurred in VMC. In Study 2, the only undetected partial understanding and the only detected but unrepaired misunderstanding occurred in VMC. There are more undetected misunderstanding problems in VMC than in FTF. These all suggest that FTF provides better chances of detecting, handling, and resolving understanding problems than does VMC.

## 11.4 Critical reflections and limitations of Study 2

As discussed earlier, understanding in interaction is difficult to approach and operationalise. A number of researchers, for example, Mustajoki (2012), Verdonik (2010), Weigand (2010), Wierzbicka (2010), and Sarangi (1994) have pointed out that counting and accounting for understanding problems and miscommunications is problematic. The present study is limited in a few ways that will be discussed below.

### 11.4.1 Identifying and quantifying understanding by means of micro-feedback and meaning repair

Study 2 in this thesis has made an attempt to quantify understanding cases by means of micro-feedback and other related responsive actions, primarily, meaning repair. In the empirical data studied, partial understanding and non-understanding have been found to be almost always revealed through vocal-verbal and gestural micro-feedback expressions that signal, respectively, *do you mean this...*, *is it this...*, and the like and *what did you say, please pardon, sorry, I do not understand, can you explain*, and the like. Gestural micro-feedback smile and multimodal micro-feedback chuckle and laughter are sometimes employed by interlocutors when they have difficulties or problems in achieving sufficient understanding, usually expressing emotions and attitudes of friendliness, politeness, uncertainty, and embarrassment.

However, sufficient understanding and misunderstanding are often related to similar micro-feedback expressions such as head nod, smile, *yeah, okay*, and the like (see Lu, 2014). It is not easy to make generalisations of how to distinguish misunderstanding from sufficient understanding. Therefore, related responsive actions other than micro-feedback, meaning repair interactions, prosodic features of vocal-verbal utterances, gestural communicative behaviours, and the local and global discourse contexts of relevance are very helpful.

This approach provides a possibility of being able to measure and compare understanding in interactions. Although it makes analysing understanding in social communication more practical, this approach is subject to an audio and video analysis and very dependent on the annotator's and coder's interpretations. What is more, it may bring up a big question for language and communication scholars of whether understanding in interaction is countable and accountable although it is most usually unfinalisable in reality.

Understanding is a complex cognitive and linguistic process. Understanding in social interaction is as important as in educational communication. If there is a way of evaluating educational understanding (e.g., oral and written exams and feedback exchanges), there may also be a way of evaluating understanding in social interaction.

Study 2 in this thesis could be regarded as one step in this direction although with various restrictions.

#### 11.4.2 Analyses from an analytical perspective

As discussed in the theory and method chapters, based on theories of social communicative activity type, meaning and implicature, contextualisation, and relevance, this thesis applies an interactional approach to investigating understanding in two empirical data sets.

Although inter-coder reliability checking, self-confrontation interviews (for Study 1), and follow-up interviews (for Study 2) were made to different extents, the analyses of this thesis (e.g., how understanding problems occur and how they are detected, handled, and resolved by the interlocutors) were primarily conducted from the analyst's point of view. This means that the analyst could misunderstand, misinterpret, over understand, and over interpret the speaker's intentions and anticipations, or the analyst could even miss understanding problems (e.g., misunderstanding) that have in fact occurred in the participants' conversations. This is a constantly discussed issue in analytical studies.

Increasing the number and variety of coders could be helpful. However, involving participants for subsequent self-confrontation interviews may or may not be as effective as expected. For instance, in Study 1 in this thesis, many participants could not answer with certainty the questions on what happened and why s/he said or did this or that. In Study 2 in this thesis, the participants were frequently found to be contradictory in their independent answers to questions about how they solved the learning assignments, what they (dis)agreed on, and what the solutions were. Attention needs to be paid to this issue.

#### 11.4.3 Sufficient understanding, acceptance, and agreement

As shown in the data, many sufficient understandings in interactions are not as obvious and significant in the discourse exchange as partial understandings and non-understandings. Very often, sufficient understanding is signalled through unimodal vocal-verbal *yeah*, *okay*, *m*, *ah*, and *yes*, unimodal gestural nods, nod, smile, up-nod, and up-nods, and multimodal *yeah+nods*, chuckle, *yeah+nod*, *m+nods*, laughter, and *okay+nods*, expressing "I have sufficiently understood what you have said and I have accepted that this is your opinion; I may agree or disagree (which I will let you know now or soon); at the moment, I am content with our interaction and please continue" (Lu, 2014).

Both acceptance of the shared information and agreement with it are based on the prerequisite of an understanding (either sufficient or insufficient) of the shared information, except for fake acceptance and agreement (associated with fake understandings), for example, for social reasons in specific contexts. In contexts where, for example, the interlocutors have conflicting goals, fake acceptance and agreement may occur. Since it is not possible to analyse such fake understandings and only the overt and manifested understandings are focused on in this study, fake acceptance and agreement are not taken into account. However, there may be different dimensions and levels in sufficient understanding. Some are shallower than others and play simpler roles in interaction discourse. This issue is not handled in depth in the present study and may need further research, perhaps taking into consideration of diverse communication contexts other than first encounters and task-solving collaboration where interlocutors are friendly and kind to each other and engaged in achieving their joint communication goals.

#### 11.4.4 University students' intercultural communication in the English lingua franca

Study 2 focuses on Swedish and Chinese university students' intercultural communication in a normal educational activity with collaborative learning tasks by using the English lingua franca.

First, students at the university level are comparatively more sophisticated than other groups of people in language use and logical reasoning. They are likely to be analytical and open-minded, and they can quickly become involved in new tasks and focus on working out solutions. A variety of participants, who have more discrepancies in knowledge and experience, which may likely cause greater difficulties in task-solving collaboration, could perhaps be of interest for the study of understanding problems in social interaction.

Second, according to a number of earlier researchers, for example, Gumperz (1982), Pride (1985), Tannen (1990), Smith (2004), Eliot (2010), Samovar et al. (2012), and Gogan et al. (2014), the context of intercultural communication should affect communication and increase understanding problems. This is, however, turned out not to be the case in any extreme way in the data focused on in this study. This is perhaps because earlier research on understanding was mainly in the form of qualitative studies, often with a theoretical focus. They did not analyse the occurrences of understanding and understanding problems from quantitative perspectives, and understanding problems were not counted in the same way as they are in the present study. Also, this result may be somewhat related to the fact that in this globalised world, people have more access to intercultural communication and they

become more proficient at it. Additionally, this study is specifically based on Swedish and Chinese participants. It will be interesting to carry out further empirical studies on understanding in communication with a greater diversity of nationalities and cultures, including both mono-cultural and intercultural communications.

Third, the participants in this study communicated in their common language, the English lingua franca. Because the participants were non-native English speakers, this data might be more interesting for research on understanding and understanding problems. However, it is not easy to identify in which way and to what extent the first language has affected the (vocal-verbal and gestural) interactions and understandings, when the speakers communicate in English as a second language. Further research is required.

The results of this study are specific to Swedish and Chinese university students' intercultural communication in a normal educational activity with collaborative learning tasks by using the English lingua franca. There is a question of how generalisable the results of the analyses can be when applying them to other communication contexts. Varying socio-cultural and lingual conventions and communication contexts can contribute to research the subject of understanding in interaction. How and how much these factors affect speech production and interpretation can be researched further.

#### 11.4.5 VMC audio and video chat

Equally important, Study 2 in this thesis focuses on VMC audio and video chat, which has been classified as a kind of synchronous spoken communication mode (Hård af Segerstad, 2002) and acknowledged as “comparable to face-to-face interaction” (Fägersten, Holmsten, & Cunningham, 2010, p. 147). Although VMC audio and video chat has been “positioned as key to facilitating meaningful teamwork activity” (Fägersten et al., 2010, p. 146), other VMC mode variants, for instance, “combining video with voice chat, text chat (i.e., instant messaging), whiteboard capabilities, and collaborative document manipulation” (Fägersten et al., 2010, p. 146) could also be of interest when investigating understanding in VMC and digital communication in the future.

### 11.5 Suggestions for future studies

Casual conversation is still important as research data for analysing micro-feedback and understanding (see earlier discussion by Verdonik, 2010). A larger set of casual dialogues (contexts) could possibly be added to the current research data.

The subject of understanding in particular understanding problems could be studied with a larger group of dynamics and variables in terms of the communication context and participants. What has been focused on in Study 2 is an interaction activity involving a complex task of jointly solving some learning assignments. The participants were classmates who were acquainted with each other. According to earlier studies, for example, Maynard and Zimmerman (1984), the acquainted and the unacquainted interlocutors show considerable differences in initiating and developing topics and mutual understandings. Interactions with the same or a similar communicative task between unacquainted participants could be of interest as the next study connecting to the present research. Also, as shown in the result from Study 2, being new to knowledge and experience is one of the most common reasons why understanding problems occurred in the data. Participants other than university students, who differ widely in education, working, and knowledge of life and experience, may likely produce more understanding problems than the participants studied in Study 2 in order to research understanding in interaction.

According to Gumperz (1982), Tannen (1990), and Samovar et al. (2012), socio-cultural backgrounds influence how people perceive, understand, and use language. Various interactions other than Swedish–Chinese intercultural communication, as focused in Study 2, could be of interest for future studies. Also, the degree to which socio-cultural conventions affect speech production and interpretation can be researched further, for example, by comparing understanding and its problems between mono-cultural and intercultural communications.

More variations of discourse contexts and participants could enrich research on understanding, which could result in further interesting findings.



# 12

## **General discussion and conclusion of the thesis**

In this closing chapter of the whole thesis, first, a review of the research purpose, questions, data, and approach will be presented. Second, the main empirical findings of the studies will be summarised. Third, contributions and implications of the thesis will be discussed. Next, critical reflections on the studies will be made and research limitations will be pointed out. Then, suggestions for the future studies will be provided. Lastly, some concluding remarks will be made.

### **12.1 Review of the research purpose, questions, data, and approach**

In this section, the research purpose, questions, data, and approach in the thesis will be reviewed.

#### **12.1.1 Research purpose**

The overarching aim of this thesis was to contribute to understanding the understanding in real-time communication by empirically investigating how understanding is signalled, detected, handled, and resolved in social interactions of varying complexity in intercultural, multimodal, and video-mediated communication situations. The analytical focuses were on micro-feedback and meaning repair. The thesis also

aimed to uncover similarities and differences in understanding between face-to-face (FTF) and video-mediated communication (VMC). Conceptual analyses of the concepts of understanding and micro-feedback were conducted when studying the inter-relations between them.

Two major empirical studies have been carried out in two activity types, where the English lingua franca is spoken. Study 1 has investigated micro-feedback in relation to understanding issues in a spontaneous communication activity in first encounters. Based on the results from Study 1, Study 2 has expanded the research and examined how understanding problems are coped with by acquainted interlocutors in relation to not only micro-feedback but also meaning repair in an educational activity with collaborative learning tasks.

### 12.1.2 Research questions

Study 1 (Chapters 5–7) has investigated three research questions. First, how are the auditory and visual modalities involved in micro-feedback expressions that are related to sufficient understanding, misunderstanding, and non-understanding? Second, what are the typical unimodal and multimodal micro-feedback expressions that signal sufficient understanding, misunderstanding, and non-understanding? Third, what specific prosodic features of vocal-verbal micro-feedback are correlated to sufficient understanding, misunderstanding, and non-understanding?

Study 2 (Chapters 8–11) has examined another three research questions regarding what understanding and understanding problems are in social communication, how understanding problems are detected, handled, and resolved in and through interaction, and what similarities and differences there are between face-to-face and video-mediated communication in the occurrence, detection, handling, and resolving of understanding and understanding problems, respectively.

### 12.1.3 Research data

The empirical data in Study 1 consists of eight audio- and video-recorded situated interactions. The participants were four Swedish and four Chinese university students, who were strangers to each other. The participants communicated in their common language, the English lingua franca. Their activity task was to get acquainted with one another.

The empirical material for Study 2 consists of 20 audio- and video-recorded interactions between ten Swedish and ten Chinese university students, who were already acquainted with each other. They spoke the language of English lingua franca. Their activity task was to solve two educational learning tasks collaboratively.

#### 12.1.4 Research approach

Study 1 has explored the phenomenon of micro-feedback based on relevance theory and contextualisation theory, with a focus on context dependency. The unit of analysis was micro-feedback, and the main classifications were based on physical properties of micro-feedback expressions.

Study 2 has examined the concept of understanding by using an interactional approach based on theories of social communicative activity type, meaning and implicature, contextualisation, and relevance. The analytical focuses were on micro-feedback and meaning repair of interactions in which utterances were assumed to be sequentially interdependent.

In the following sections, the empirical findings of the studies will be summarised.

### 12.2 Summary of the empirical findings

With a criterion of whether the information is understood sufficiently and correctly in relation to what is required to continue the interaction and what can be inferred about the interlocutor's intention and anticipation, the different types of understanding—(overt) sufficient understanding, partial understanding, misunderstanding, and non-understanding—have been identified. Specific patterns of micro-feedback in terms of modality and prosody in relation to understanding and features of understanding in terms of its occurrence, detection, handling, and resolving varying between FTF and VMC, have been identified. In the data collected for this thesis, sufficient understanding is most common in spontaneous communication, the second most common is partial understanding and non-understanding, and the least common is misunderstanding.

The following sections will synthesise the empirical findings in Study 1 and Study 2 to address the research questions in relation to each type of understanding.

#### 12.2.1 Sufficient understanding

Sufficient understanding occurs when it is enough to serve the current practical purposes of information sharing and sense-making, no matter how partially it is shared. It is usually exhibited in the speech acts of declaring and persuading. The information presented is understood in a way that is correct for current purposes in relation to what is intended and anticipated, no matter how much is in fact correct. Interlocutors are content with understanding one another and feel good enough to proceed further in the interaction.

Sufficient understanding is more related to unimodal micro-feedback than multimodal micro-feedback. Sufficient understanding is frequently signalled by unimodal vocal-verbal *yeah*, *okay*, *m*, *ah*, and *yes*, unimodal gestural nods, nod, smile, up-nod, and up-nods, and multimodal *yeah+nods*, chuckle, *yeah+nod*, *m+nods*, laughter, and *okay+nods* (top five for each). The vocal-verbal micro-feedback is associated with a small pitch range, a flat or falling pitch contour, and a short or a medium duration. Unimodal head movements exclusively signal sufficient understanding.

### 12.2.2 Partial understanding

Partial understanding is one form of insufficient understanding or understanding problem. It occurs when one cannot sufficiently, and only partially, make sense or share the meaning of the information presented as intended or anticipated. Partial understanding is almost always detected and usually handled by means of other-initiated repair questions, which suggest, repeat, or paraphrase what has been partially understood and seek further confirmation, clarification, explanation, or specification. This is a developing stage towards achieving an eventual sufficient understanding. Partial understanding is only observed in task-solving interactions in Study 2 and not in first encounters in Study 1.

Partial understanding is usually revealed by micro-feedback expressions such as *do you mean this...*, *is it this...*, and the like. Also, gestural micro-feedback smile and multimodal micro-feedback chuckle and laughter which primarily express emotions of friendliness, politeness, uncertainty, and embarrassment are sometimes signals of partial understanding. The detected partial understanding problems are always repaired and thus promoted to sufficient understanding.

### 12.2.3 Non-understanding

Non-understanding is another form of insufficient understanding or understanding problem. It occurs when the information presented is not understood at all for reasons such as lack of access to the referents or concepts themselves or some background knowledge of relevance. It cannot serve the current practical purposes of sharing and making sense of the information presented. It does not correspond to what is intended and anticipated. Normally, non-understanding is manifested in the speech act of questioning. It is always detected and most often coped with by the interlocutors, and it usually leads to sufficient understanding in the end.

Non-understanding is often handled by means of other-initiated repair questions, with the aim of eliciting further clarification and explanation. Unimodal vocal-

verbal micro-feedback such as *sorry* and *what do you mean*, unimodal gestural micro-feedback eyebrow rise and gaze at, and multimodal micro-feedback eyebrow rise or frown, head forward, gaze sideways, or gaze at the speakers plus *sorry*, *what*, or *huh*, and sometimes chuckle and smile are usually involved, signalling *what did you say*, *please pardon*, *sorry*, *I do not understand*, *can you explain*, and the like. In these expressions, the vocal-verbal micro-feedback component is associated with a small pitch range, a rising pitch contour, and a short duration. It is more frequently signalled through multimodal micro-feedback than unimodal micro-feedback. Non-understanding problems are always detected but not always repaired, often because the interlocutor wants to ignore them or wants to repair them but fails, and thus abandons any further repairing.

#### 12.2.4 Misunderstanding

Misunderstanding is also a form of insufficient understanding or understanding problem. It only occurs when the recipient interlocutor comes up with an interpretation of the communicated information but in an incorrect way or deviates from what is intended or anticipated. Although it can perhaps serve the current practical purposes of exchanging information and carrying on the conversation, it does not always lead to sufficient understanding. Instead, it can sometimes cause further misunderstandings. Misunderstanding is not always noticed. Usually, the interlocutors just carry on their interaction without being aware of it. Misunderstanding is more observable in task-solving interactions in Study 2 than in first encounters in Study 1.

Detected misunderstandings are handled more often by other-initiated meaning repairs than self-initiated ones, and they are almost always repaired and corrected and thus promoted to sufficient understanding. Compared to partial understanding and non-understanding, misunderstanding is found to play a more special role in sense-making and information sharing. It is least related to micro-feedback, and it is most difficult to observe in interaction. As a matter of fact, misunderstanding is very often undetected, and the undetected misunderstanding is constructive for interaction but often unconstructive for the intended sense-making and information sharing. Misunderstanding normally occurs when *yeah* and nod are used with hesitation and uncertainty and *yeah* is associated with a small pitch range, a falling pitch contour, and a short duration. It is more frequently related to multimodal than unimodal micro-feedback. Unimodal gestures are not involved in misunderstanding at all. Also, a local or a global context can help to identify misunderstanding. Additionally, when the information presented is repeated, paraphrased, or responded to with unanticipated actions, a misunderstanding has likely occurred.

### 12.2.5 Generalisation

The analyses were made of a particular set of data, in which participants have joint and harmonious communication goals. The results of this material can indicate certain general trends that may hold for other but not all social interactions.

This thesis confirms that micro-feedback is one evidence of showing understanding (at least shallow understanding) (see also Clark & Schaefer, 1989; McConnell, 1993; Ryan & Conover, 2004; Linell, 2011; Boud & Molloy, 2013; Bertrand & Goujon, 2017). Both modality and prosody of micro-feedback play significant and systematic roles in communication and understanding (see also Grice, 1989; Couper-Kuhlen & Selting, 1996; Ward & Tsukahara, 2000; House, 2006; Mitchell & Ross, 2013; Goodwin, 2018). When studying understanding in interaction, the social communicative activity type (e.g., Wittgenstein, 1958; Allwood, 1976; Levinson, 1979), the implication of intended meaning and anticipated reaction (Grice, 1975), the context of interaction (Gumperz, 1982), and the discourse of relevance (Sperber & Wilson, 1986) should be all taken into account.

Features of modality and prosody of micro-feedback may shed light on indicating different understandings. For example, sufficient understanding is more frequently related to unimodal micro-feedback than to multimodal micro-feedback, but understanding problems are more frequently related to multimodal micro-feedback than to unimodal micro-feedback. Unimodal head movements exclusively signal sufficient understanding, and unimodal gestures are not involved in misunderstanding at all. Gestural micro-feedback eyebrow rise or frown, head forward, gaze sideways (from) or gaze at (the other interlocutor), and sometimes smile, as well as multimodal micro-feedback chuckle and laughter can be indicators of understanding problems. Partial understanding is usually revealed by micro-feedback such as *do you mean this...*, *is it this...*, and the like. Non-understanding is often expressed through micro-feedback that signals *what did you say*, *please pardon*, *sorry*, *I do not understand*, *can you explain*, and the like. Partial understanding and non-understanding are usually associated with eliciting questions in a rising pitch contour. Misunderstanding is difficult to distinguish from sufficient understanding, because both involve similar micro-feedback expressions such as head nod, *yeah*, and *okay* associated with a small pitch range and a falling pitch contour. A local or a global context can help to identify misunderstanding. Interlocutors are warned to be aware of misunderstanding when information is repeated, paraphrased, or responded to with unanticipated actions. Micro-feedback, for example, nod and *yeah* associated with hesitation and uncertainty may also indicate misunderstanding.

Furthermore, as suggested in the analysis, technology does not decrease the probability of sufficient understanding or increase that of understanding problems

(see also Anderson, 2006). In this thesis, it has been found that it is slightly easier to detect and repair understanding problems in FTF than in VMC. Especially as regards misunderstanding, it is difficult to observe misunderstanding problems in interaction and even more difficult in VMC than in FTF. Understanding problems are more commonly handled by other-initiated but self-performed meaning repairs, which confirms Schegloff, Jefferson, and Sacks' (1977) findings.

In general, in order to understand understanding in social interaction, inter-dependency between micro-feedback (both gesture and prosody), understanding, and context has to be taken into account (see also Lindwall, 2008; Linell, 2009; Nadeu & Prieto, 2011; Hindmarsh et al., 2011; Finkbeiner et al., 2012; Goodwin, 2018). Micro-feedback does not provide all the sufficient conditions for identifying and evaluating understanding in interaction, but it does provide some necessary conditions and assistance for this (see also Lu, 2014; Bertrand & Goujon, 2017). However, with the contextualised and interactional accounts of inference and relevance, primarily the context dependency consideration, micro-feedback and its modality and prosody features can provide the analyst and the interlocutor with more information about how the communicated message has been understood (see also Lu, 2014).

## 12.3 Contributions and implications of this thesis

This thesis makes theoretical and practical contributions. This section will first present the implications of the research findings and how these findings may have repercussions on the existing theories of micro-feedback in relation to understanding in interaction. Following that, how this thesis contributes to practical knowledge and use will also be addressed.

### 12.3.1 Theoretical implications

From a theoretical perspective, I will discuss the concepts of micro-feedback and understanding and their operationalisation, how empirical findings in this thesis are in line with or contradict some earlier theories, and how the present research work could possibly influence future studies on this subject.

#### Terminology: micro-feedback versus minimal feedback and feedback

As presented earlier, *micro-feedback* has been widely named as *backchannel* (Yngve, 1970), *minimal feedback* (Fishman, 1978), *continuer* (Schegloff, 1982), *feedback* (Allwood, 1993), and so on. From a semantic perspective, the terms Fishman's *minimal feedback* (1978) and Allwood's *feedback* (1993) are closer than others to *micro-feedback* used in this thesis.

*Minimal feedback* (Fishman, 1978, p. 402), however, only refers to the minimal responses that are short words such as *yeah*, *umm*, and *huh*. It excludes responses that consist of long phrases or even sentences like *yeah yeah yeah you're right* and *sorry what did you say*, and it excludes gestures. *Micro-feedback*, however, takes both short and somewhat longer linguistic responses into account, and it also includes the *micro-feedback* gestures. *Minimal feedback* and *micro-feedback* are similar semantically but different conceptually and empirically.

The term *feedback* is used in a wide sense in the field of language and communication. For instance, it is used to refer to communication responses (Wood, 2011), or comment and evaluation (Mahboob & Knight, 2010), or some particular linguistic device that has certain communicative functions (Nivre et al., 1992), for example, *yeah yeah yeah* or head nod that signals “I hear and understand what you have just said”. This thesis attempts to reduce or avoid this terminological ambiguity. Also, it is preferable to highlight the pragmatic feature of being small and unobtrusive in the semantic definition. Accordingly, the term *micro-feedback* is used in this thesis.

### Definition and operationalisation of micro-feedback

This thesis work has adopted Nivre et al.'s (1992) and Allwood's (1993) conceptualisation of *feedback*, referring to the unobtrusive (i.e., without taking the turn or making any substantial contribution to the content or discourse dimension of the interaction) “verbal and nonverbal communicative expressions that are used to give and elicit information” (p. 1) about the continuation of the interaction, the perception and understanding of the information communicated and the attitudinal and emotional reactions to the perceived and understood information (i.e., CPUE/A).

The research was mainly based on Kopp et al.'s (2007) and Grammer et al.'s (2008) frameworks for analysing *feedback*, in particular in terms of its basic communicative functions (i.e., CPUE/A) and its unimodal and multimodal properties (i.e., how auditory and visual modalities are involved).

Developing from there, in addition to Nivre et al.'s definition of *feedback*, *micro-feedback* in this thesis has been also defined with the following features: (1) *micro-feedback* has no independent referential or semantic meaning but is very much dependent on the communication context; (2) it occurs during or after the other speaker's talk, preferably at “response points” (Linell, 2011, pp. 266–270), usually at the beginning of a responsive communication contribution, which includes both

spoken utterances and gestural behaviours<sup>59</sup>; (3) besides CPUE/A, *micro-feedback*, also functions as a connector between the adjacent communication contributions; (4) it sometimes expresses positive and negative evaluative opinions (e.g., agreement and disagreement).

In addition to the frameworks for analysing *feedback* by Kopp, et al. (2007) and Grammer et al. (2008), this thesis also investigates *micro-feedback* in terms of both modality (auditory and visual) and prosody at the same time. In particular, in terms of prosody which is in particular related to understanding, the studied phenomenon *micro-feedback* has been measured with respect to pitch (i.e., pitch values, pitch range values and types, and pitch contours), duration (i.e., duration values and types), and the inter-relations between them in this thesis, which has not been done in earlier research on *feedback*. The result is an enriched definition and operationalisation of *micro-feedback*.

### The concept of understanding and its classification: in relation to meaning and implicature

As Allwood (1986) claimed, linguistic communication can be reducible to two main processes, namely the understanding and the production of communicative behaviours. In all communication, a fundamental problem has to do with understanding (cf. Allwood, 2015). The general purpose of communication is to arrive at an understanding. As soon as people perceive some information, they start making sense of it and understanding it with regard to values of relevance (see *relevance theory* in Sperber & Wilson, 1986; Zlatev, 2009) to the implication of intended meaning and anticipated reaction (see *meaning and implicature theory* in Grice, 1975). However, it is not just a matter of understanding the topic or content that is talked about, but also of understanding the other's (possibly discrepant) understandings.

Many researchers believe that understanding in interaction is always pragmatic and situated understanding, for example, Garfinkel (1967), Zaefferer (1977), Allwood (1986), Bakhtin (1986), Dascal and Berenstein (1987), Taylor (1992), Weigand (1999), Weizman (1999), Linell (2009), and Lindwall and Lymer (2011). Understanding the information presented means to infer and understand the communicative purpose and the anticipated reaction the information presented is related to. This mainly involves an interaction between the linguistically conventional meaning

---

<sup>59</sup> It should be noted that some of these authors include feedback elicitation. However I have not studied the giving and eliciting functions of *micro-feedback* specifically but only focused on the function of understanding of *micro-feedback* in this thesis. Also, I have not explored pauses and delays before next turn or utterance, which can be seen as a kind of (lack of) *micro-feedback*.

and the speaker's intended and exhibited meaning as well as other relevant contextual conditions (see literature reviews in Chapter 2).

Based on Allwood's (1986), Zaefferer's (1977), and Weigand's (1999) suggestions for classifying understanding and Linell's (2009) suggestion of partially shared but sufficient understanding as well as the fact that understanding is closely related to *response* and *anticipation* (see also Bakhtin, 1986, p. 69), the criterion for categorising understanding is established based on whether the information is understood *sufficiently* well and *correctly* in relation to what is required to continue the conversation, including what can be inferred about the speaker's intention and anticipation (Grice, 1975). Analytical results have shown that there are four types of understandings in the empirical data, which include sufficient understanding, misunderstanding, partial understanding, and non-understanding.

The concepts of *sufficient understanding*, *misunderstanding*, *partial understanding*, and *non-understanding* have been used, respectively, by Garfinkel (1967), Bakhtin (1986), and Taylor (1992). See Linell (2009, p. 222) for *sufficient understanding*, Weigand (1999, p. 763) and Allwood (1986, p. 20) for *misunderstanding*, Allwood (1986, p. 20) for *partial understanding*, and Zaefferer (1977, p. 337) and Weigand (1999, p. 765) for *non-understanding*. They vary to some extent from what have been defined in this thesis, mainly because of semantic and pragmatic reasons (see Chapter 2). To set up a criterion for evaluating understanding on a micro level can be further studied in the future. This thesis places the focus on these four types as a starting point.

## Intercultural communication and understanding

Earlier studies, for example, Gumperz (1982), Tannen (1990), Samovar et al. (2012), and Allwood (2015), have suggested that in a joint communication activity, people who have different cultural and language backgrounds probably have more problems and difficulties in understanding than those who have the same relevant backgrounds. Because socio-cultural conventions affect many or all levels of speech production and interpretation (Gumperz, 1982; Pride, 1985; Tannen, 1990; Samovar et al., 2004; Eliot, 2010; McDaniel, 2012), cultural differences affect communication and increase understanding problems (Gogan et al., 2014).

Intercultural communication has likely higher risks of a lack of understanding and misunderstanding (Allwood, 2015; Lindström, 2008). This is also why the focus of the thesis was on intercultural interactions. However, the results of the empirical studies in this thesis show that there are not so many understanding problems as expected. For example, out of 1288 understanding cases studied in Study 1 of

strangers becoming acquainted with one another, there are only 32 instances of understanding problems, which include 9 cases of misunderstanding and 23 cases of non-understanding. Although Study 2, with a more complex communication task of jointly solving learning assignments, has produced more understanding problems than Study 1, there are 153 cases of understanding problems identified in Study 2, including 21 misunderstandings, 69 partial understandings, and 63 non-understandings compared to a frequency of 2204 understanding cases in all. Maybe the participants studied did not have very many understanding problems because that they were similar in terms of age and generation, and they had similar education and learning levels, experiences, and competences. In addition, it may have something to do with the fact that normally interlocutors are friendly and kind to each other. Thus, it is not surprising that in normal communication situations people have more sufficient understandings than insufficient ones. It may be also because “communication failures” in intercultural encounters are normal as they take place between good friends and relatives (cf. Mustajoki, 2012, p. 217), which somewhat contradicts the earlier theoretical assumptions about understanding problems in intercultural communication. This may need further empirical investigation. This suggests further studies with more variations in communication participants in similar or more various communication activities.

#### Is misunderstanding ubiquitous? How to identify misunderstanding?

The view that misunderstanding is ubiquitous in conversation, which was suggested by Fraser (1993) and Dascal (1999), is not supported by the present analysis of understanding. In empirical Study 1, only 9 out of 1288 instances of micro-feedback expressions related to understanding have been identified as misunderstanding occurrences. In empirical Study 2, there are 21 misunderstanding cases identified from a total of 2204 understanding instances. Thus, in this thesis, misunderstanding does not occur often. In fact, compared to sufficient understanding and the other two types of understanding problems (i.e., partial understanding and non-understanding), misunderstanding occurs least frequently. Therefore, misunderstanding does not seem as ubiquitous as it has been assumed to be.

One issue is how to identify misunderstanding. According to the empirical results of Study 1, both sufficient understanding and misunderstanding involve similar micro-feedback expressions such as head nod, *yeah*, and *okay*, which are associated with a small pitch range, a falling pitch contour, and a short duration. Consequently, it is especially difficult to distinguish misunderstanding from sufficient understanding in interaction. For this reason, the observable features of misunderstanding in relation to micro-feedback and discourse context should be paid more attention.

First, misunderstanding usually involves *yeah* or a repeated noun phrase of the perceived information and nod, which are sometimes associated with emotions and attitudes of hesitation and uncertainty. Second, when misunderstanding occurs, unimodal gestures are not involved. Third, misunderstanding is usually associated with a small pitch range, a short duration, and a falling pitch contour (see Table 7.1), and thus it is negatively associated with the medium and long durations and the flat and falling pitch contours. Fourth, a local or a global context can help to identify misunderstanding. Fifth, when information is repeated, paraphrased, or responded to with unanticipated actions, a misunderstanding may have occurred. Next, it is not always corrected or even pointed out by the interlocutors in the interaction.

If both misunderstanding and non-understanding are regarded as difficulties or problems in understanding, this finding does not support Clark and Schaefer's (1989) statement that the interlocutors repair any trouble they encounter with each other's utterance. Instead, this finding is consistent with what Weigand (1999) and Verdonik (2010) claimed, namely that misunderstanding is not always corrected by the speakers. This might be because of, for example, the speaker's interest in the topic, the necessity of clarifying the misunderstood information, and the time allowed. In these cases, some kind of vagueness or ambiguity may occur. However, it may be also because the interlocutors are in fact not aware of it. As the data have shown, one misunderstanding can result in further misunderstandings between the participants back and forth throughout the conversation, without the participants being aware of them.

Nevertheless, in order to identify misunderstanding, whether from an analyst's perspective or an interlocutor's perspective, a richer context with more information on the involved prosody and gesture should always be taken into account. That is, both modality and prosody of speech (e.g., micro-feedback in this thesis) as well as other responsive communication behaviours (e.g., meaning repair) and the relevant discourse context can assist in identifying misunderstanding. This underlines the importance of the interactional approach to studying understanding and context dependency for analysing communication.

### Interactional approach to understanding: relevance and contextualisation

There are a number of influential approaches to studying understanding, as has been primarily discussed in Chapter 2. The interactional approach (Sacks et al., 1974; Lynch, 2011) is adopted in this thesis, as understanding is viewed as an interactive and situated phenomenon shown by the use of language and the construction of discourse. Understanding is something that is visible in discourse, and it is negotiated by the interlocutors through speech and context (e.g., in part contextualisation, see

Gumperz, 1982; Bauman & Briggs, 1990; Tannen, 1993; Couper-Kuhlen, 2001). For instance, with analyses of interactional context in which some misunderstandings (as discussed above) evolve and are sometimes repaired, an interactional approach is highlighted in order to study understanding and in particular misunderstanding in social interaction.

In this thesis, the operationalisation of understanding was conducted based on Garfinkel's (1967), Zaefferer's (1977), Allwood's (1986), Weigand's (1999), and Linell's (2009) ontological assumptions and frameworks for classifying understanding (as presented earlier in Chapter 2). All of them touched upon the relevance theory (Sperber & Wilson, 1986) and contextualisation theory (Gumperz, 1982) to varying extents, namely the discourse of relevance and the context of interaction in this thesis.

*Relevance* has been taken as a prerequisite of understanding, and understanding is a necessary requirement for a speaker to make a relevant contribution to the cooperative communication activity. Relevance is a better indicator of understanding than a linguistic means such as micro-feedback itself in social interaction and also in the interactions between utterances. This is comparable to the approach of *contextualisation*, which stresses the importance of communication context in indexing or evoking interpretive schemas or frames within which inferential understanding is achieved (Gumperz, 1982; Tannen, 1993) and the context effect on language use and discourse construction, which at the same time also reflect the relevant aspects of the interactional context (Bauman & Briggs, 1990; Couper-Kuhlen, 2001). Interpretation and determination of what people are doing in the interaction has to be based on contextual accounts. Although turn taking, sequencing, and interaction between utterances or discourse coherence are not focused on in this thesis, the analysis of micro-feedback and understanding still relies very much on the relevant speech and its accompanying gestures and prosody as well as the relevant context as a whole.

### Interdependency between context, micro-feedback, and understanding

As discussed earlier, micro-feedback is one way of showing understanding in social interaction and it assists in the management of communication. In this thesis, the data from first encounters and task-solving collaboration have generated a certain number of micro-feedback properties and the related understanding issues. Specific features and patterns of micro-feedback in terms of its modality and prosody have been found to be associated with sufficient understanding, misunderstanding, and non-understanding. This confirms the claims of, for instance, Maynard and Zimmerman (1984), Svennevig (1999), Campbell (2007), Navarretta and Paggio (2013), and Lu (2014) that micro-feedback has something to do with the communication of

understanding. Although the findings in this thesis do not provide all the sufficient conditions for identifying and evaluating understanding in communication, they do provide some necessary conditions and assistance for both the analyst and the interlocutors to obtain more insights into how the communicated information has been understood. To put it differently, micro-feedback can be used as signals of only some kind of understandings. In order to identify and evaluate understanding in social interaction, in addition to the vocal-verbal speech (of the turns in their entirety) together with its various prosodic aspects and its relevant gestural information, relevance and contextualisation should be taken into account (see Lu, 2014). There is no escaping the fact that a fairly high degree of contextual interpretation (i.e., contextualisation) is present in such an empirical study of micro-feedback in relation to understanding.

Obviously, the interdependency between context, micro-feedback (both modality and prosody), and understanding cannot be ignored. On the one hand, context links the vocal-verbal, prosodic, and gestural micro-feedback behaviours (as discussed above); on the other hand, these linguistic behaviours also help construct the context. None of them can be separated from the context or even from each other, when talking about sense-making and understanding in communication. None of them in themselves signal or determine meaning and understanding, but they jointly contribute to the process of contextualisation and interaction. As found in the empirical studies in this thesis, vocal-verbal and gestural aspects of multimodal micro-feedback have shown their interdependent and complementary relationships. Various prosodic aspects of the vocal-verbal micro-feedback and the accompanying gestures of relevance have added to and changed the semantic meanings of speech, thus adding and changing the common ground for sense-making and understanding. Even within the prosody of micro-feedback, pitch contour, duration type, and pitch range type have been found to be inter-associated with one another in relation to the different understandings.

In general, the temporal and situational context is important when determining micro-feedback and understanding in social interaction, and micro-feedback functions in relation to understanding and helps construct the context and manage the interaction.

## Social communicative activity type: expectation and anticipation of language use and understanding

Results show that partial understanding is not found in Study 1 of first encounters but only in Study 2 of task-solving interactions. Besides, more understanding problems occurred in Study 2 than in Study 1. Also, misunderstanding is more observable in task-solving interactions than first encounters.

Explanations for these results can be drawn from the theories of social communicative activity type, for example, Wittgenstein's (1958) *language games*, Allwood's (1976) *behaviour, situation, and meaning types*, Levinson's (1979) *activity type*, Allwood's (2013) *activity-based communication analysis (ACA)*, and Linell's (2010) *communicative activity type (CAT)*. Research by, for example, Levinson (1979), Goffman (1974), and Gumperz (1982) has pointed out that activity type plays an important role in language use and understanding. On the one hand, social activity type constrains what will count as allowable contributions to the communication; on the other hand, social activity type helps to determine what kinds of inferences and how they will be made from what is communicated (see also Linell, 2010).

Because there are often special relations between what is said and what is done with inference to what can be said and what can be done, people usually have specific expectations and anticipations regarding interactions in relation to specific social activities which very much influence their language use and understanding. If understanding is acknowledged as the general purpose and fundamental problem of communication (Allwood, 2015), social activity type is the premise of understanding and communication. Different social communicative activities can contribute to research on the subject of understanding in interaction.

## Video communication technology's impact on understanding

The study (see details in Chapter 10) has shown that video communication technology does not have any impact on people's understanding in interactions: it neither decreases sufficient understanding nor increases understanding problems. This does not seem to support the claims that people often have more trouble in understanding others' responses in VMC than in FTF (Kiesler & Sproull, 1992). Neither does it support the contention that people are more likely to experience ambiguity and problems of understanding in VMC (Shin, 2005). On the contrary, the results seem to be in agreement with Anderson's (2006) finding that video mediating technology has little effect on the content of the interaction and the understanding of the interlocutors.

By using the same research method in the analysed data, the frequencies of micro-feedback-related understandings are found to be fairly close in VMC and FTF,

with 1113 and 1091 occurrences, respectively. This suggests that VMC and FTF have similar conditions for understanding and micro-feedback. It does not seem to support the claim by Paul et al. (2004) and Friedman and Currall (2003) that micro-feedback is often unclear and infrequent in VMC.

However, video communication technology has an impact on the detection, handling, and resolving of understanding problems in interactions. The thesis has found that it is more difficult to detect and repair the understanding problems, resolve them and achieve sufficient understanding, and accomplish the communication goals in VMC than in FTF. These empirical results seem to be in one way or another consistent with earlier findings that VMC is more constrained when it comes to finding common ground and achieving understanding than FTF (Olson & Olson, 2000; Clark, 1996; Clark & Brennan, 1991) and that interlocutors in VMC seem to have more difficulties in achieving communication goals (cf. Anderson, 2006) or have to work harder to achieve them (Whittaker, 2003). This might be explained by earlier theories that not all the sources of information available in FTF are transmitted to VMC (cf. Anderson, 2006) and that VMC's communication channel is relatively narrow and its media richness is comparatively low (Sins et al., 2011). Also, the higher "social presence" in FTF (Short et al., 1976; cf. Anderson, 2006) may bring about higher interactivity and interdependency between interlocutors (Stone & Posey, 2008; Kiesler & Sproull, 1992) so the interlocutors can more easily sense when understanding deviates from intention.

The thesis has a few findings regarding misunderstanding in FTF versus VMC. First, results show that it is more difficult to observe misunderstanding problems in VMC than in FTF. This is perhaps because VMC restricts the exchange of auditory and visual communication cues for prosody and gesture, which normally help people to regulate interaction, perceive, express, and comprehend information (Driskell & Radtke, 2003) and monitor feedback from others (Straus, 1997) and which have been recognised as important as a discourse context to help distinguish misunderstanding from sufficient understanding (Lu, 2014). Second, this study has also found that all the detected misunderstandings were tackled or resolved with other-initiated repairs in VMC but resolved with self-initiated repairs in FTF. This seems to support the view that FTF has higher interdependency and interactivity between interlocutors than does VMC (Stone & Posey, 2008; Kiesler & Sproull, 1992). Third, all the detected misunderstanding problems are corrected and promoted to sufficient understanding in FTF but not so in VMC. This result seems to be in agreement with the earlier studies by Olson and Olson (2000), Clark (1996), and Clark and Brennan (1991) in that VMC is more constrained to finding common ground and achieving understanding than FTF. This result is also in line with Thompson and Covert's

study (2006), in which they find communication technologies often lead to difficulties in detecting and correcting misunderstandings.

### 12.3.2 Practical implications

From a practical perspective, how the empirical findings of this thesis may affect communication and technology practice will be discussed.

#### Importance of visual modality and auditory prosody in communication technology applications

The thesis has found that the visual modality (i.e., gesture) plays an important role in constituting micro-feedback and conveying understanding in human communication. The results show that gestures are involved in around 74% of all the micro-feedback expressions that are related to understandings (based on Table 5.1). Furthermore, in the thesis, these micro-feedback gestures are almost entirely limited to the head region in the form of head movements and facial expressions. Hand, posture, and shoulder movements rarely occur as part of micro-feedback (bodily gestures are resources evidently used by speakers rather than listeners). Also, among all the micro-feedback head movements, unimodal head movements have been found to exclusively signal sufficient understanding. Smile, chuckle, and laughter associated with emotions of friendliness, politeness, uncertainty, and embarrassment can be indicators of understanding problems of partial understanding and non-understanding. Head nod, *yeah*, *okay*, and such typical signals for sufficient understanding can be indicators of misunderstanding, especially when they are associated with emotions in the form of uncertainty and hesitation. Emotional and attitudinal reactions in communication are mainly uncovered by gestural and prosodic features.

Based on these and other empirical findings (presented earlier in this thesis) of how micro-feedback and understanding are associated in terms of the modality and prosody, suggestions and contributions can be made to develop guidelines for the design of communication technology applications, such as systems for speech, gesture, and understanding recognition. Also, a basis for developing practical strategies can be obtained for simulating human spontaneous communication behaviours in artificial systems and virtual agents. The visual parts of these systems (as mentioned above), such as graphical display and motion capture, could possibly be limited to the head region of the participants or the artificial systems (e.g., virtual agents) without compromising the signalling or communication of understanding in the interaction. Similarly, these findings could contribute to the practice of technology-mediated communication, such as business consulting and cooperation, educational and

learning communication, video conferencing, virtual agents' animation and human-computer interaction involving both auditory and visual modalities.

### Patterns of micro-feedback and understanding in social interactions with harmonious goals

As presented in the empirical studies, Study 1 in this thesis has investigated micro-feedback in relation to understanding in spontaneous communication, and has identified a number of specific patterns and features of modality and prosody with respect to how micro-feedback and understanding are related. Study 2 in this thesis has re-conceptualised understanding and its problems in FTF and VMC interactions, and has identified the possible patterns of how problems of partial understanding, misunderstanding, and non-understanding are detected, handled, and resolved in practice.

Because the empirical data involves culture (i.e., Swedish and Chinese), gender (i.e., male vs. female), and individual differences (e.g., the participants each interacted once with one male and one female) as well as variations in micro-feedback expression and vowel length (e.g., micro-feedback word, phrase, and sentence), this thesis did not find any general patterns of, for example, the values of pitch, pitch range, and duration. However, the specific patterns and features of modality and prosody of the micro-feedback identified in relation to understanding in this thesis, which are generalisable and have been presented earlier, may suggest some study-specific and possible universal patterns for normal (day-to-day) social interactions in particular where interlocutors have harmonious communication goals.

Hopefully, these specific and possibly universal patterns (or general trends) of micro-feedback and other linguistic means (e.g., meaning repair) regarding understanding can be further explored and also provide communication and technology researchers with some useful information for their design work of, for example, system simulation and virtual interaction. Furthermore, these findings could help educators and learners to acquire more knowledge of or insights into whether the information communicated has been sufficiently understood, partially understood, misunderstood, or not understood through or in relation to micro-feedback and the like in their educational communication and learning collaboration.

### Some implications of cultural differences between Swedish and Chinese

Besides some specific and possibly universal patterns, there are some differences between the Swedish and the Chinese participants when using micro-feedback in relation to understanding. As regards unimodal vocal-verbal micro-feedback, only the Chinese use them in relation to misunderstanding and non-understanding while the

Swedes use them only to express sufficient understanding. The Swedes tend to use more single and repeated up-nod(s) in relation to understanding than the Chinese (see also Lu & Allwood, 2011; Navarretta et al., 2012). The Chinese tend to use more gaze movements than the Swedes (see also Lu & Allwood, 2011).

These findings can be applied directly to intercultural communication, especially first encounters between Swedish and Chinese. Furthermore, they can also help people to acquire better knowledge about how understanding is shown in communication through or in relation to micro-feedback. As a result, self-recognition and awareness of being understood, misunderstood, and not understood in daily communication activities can be enhanced and more effective and smooth interactions can eventually be achieved. Further research would be needed to strengthen and expand the findings beyond the cultural, language, and communication activity limitations of this study.

## **12.4 Critical reflections and limitations of the studies**

As discussed earlier, the design and methods of a study are very dependent on the specific goals of the researchers and other factors such as technological and time resources. This thesis has been conducted with certain limitations.

### **12.4.1 Regarding social communicative activity, participant, and dyadic dialogue**

This thesis consists of two studies of two different social communicative activities. Study 1 is based on FTF first encounters, with a simple task of getting acquainted with one another. Study 2 focuses on a complex context of FTF and VMC acquaintances' educational task-solving collaboration, with a complex task of jointly solving some learning assignments.

Unacquainted people in Study 1 apparently employed numerous questions and answers to elicit and give micro-feedback in their conversations. However, they did not manifest many misunderstandings or non-understandings and did not reveal any partial understanding. Perhaps, first encounters are restricted to studying understanding problems in the sense that unacquainted people normally do not have a need for understanding of deeper issues than facts and they normally do not want to reveal difficulties or problems in understanding in face-to-face spontaneous dyadic conversations. Also, as mentioned by Svennevig (1999) in his study, unacquainted people engage in interrogative behaviour (i.e., asking questions to acquire information about the "other") and self-disclosure (i.e., providing information about the "self"). The shared information is mainly about factual data about who they are, what

they are doing, where they live, and how old they are, which apparently neither lead to any partial understanding of the shared information nor generate much misunderstanding or non-understanding.

Although both Svennevig (1999) and the present study analysed interactions in first encounters, the studies differ when it comes to the social and interpersonal relationship between the unacquainted participants. That is, the participants in Svennevig (1999) had a long-term social and interpersonal relationship, for they were going to take a course jointly afterwards, whereas the participants in the present study did not have a presupposed (unless voluntary) long-term relationship in any joint social activity after their participation in the research project. Perhaps, in addition to Svennevig's (1999, pp. 1–6, see also Section 4.2.2) three constitutive components for interpersonal relationships: (1) solidarity, involving a set of mutual rights and obligations; (2) familiarity, involving mutual knowledge of personal information; (3) affect, involving mutual liking (or disliking), this aspect of range or duration involving mutual prospect of the interpersonal relationship could also be taken into account when studying social communicative activity.

Furthermore, acquainted people in Study 2 revealed more variety in and a larger number of understanding problems in their task-solving collaboration. Partial understanding emerged as a nuance of the classifications of understanding presented in Study 1. However, misunderstanding still occurred with a low frequency compared to partial understanding and non-understanding. Perhaps acquaintances have a higher degree of familiarity with each other and thus at lower risk of misunderstanding. This does not seem to be supported by Study 1, since there not many misunderstandings occurred between strangers either. Alternatively, perhaps misunderstanding is not ubiquitous in reality, and it does not occur as often as other understanding problems in social interactions regardless of whether the participants are acquainted or unacquainted.

However, the results may have something to do with the size and genre of the interaction. First, the participants in both studies are Swedish and Chinese students studying at universities in Sweden. They had passed certain entrance tests and satisfied higher education requirements before they were enrolled in university level studies in Sweden. The participants probably have more shared knowledge and experience than those who do not have similar education levels, which may also decrease the incidence of understanding problems. Second, the conversations in both studies are dyadic dialogues between two participants. This may have restricted the occurrences of understanding problems. Maybe increasing the number of participants could generate more understanding problems, because of the increased bias in shared information (cf. Anderson, 2006).

In addition, results have shown that there is no difference in the occurrences of sufficient understanding and understanding problems between FTF and VMC in the data studied. It could be that the limited communication channel in VMC is not that crucial, which may be because, for example, it is easier for the acquainted people to communicate and understand one another. Another explanation could be that maybe VMC does not differ from FTF as much as before, since ICT and digitalised communication have become more common in all kinds of social and personal interactions and people may have already got used to it. Alternatively, it could be that video mediating technology does not affect people's understanding in practice, and people's understanding abilities and properties stay the same or similar with or without the influence of technology.

These two specific contexts of Swedish–Chinese FTF and VMC intercultural communication only provide us with a starting point for the analysis of micro-feedback and understanding in social interaction and between FTF and VMC. Possibly, a larger group of dynamics and variables in terms of communication context, as mentioned above, with other cultures and languages could also shed more light on understanding issues in spontaneous communication.

#### 12.4.2 Different levels of micro-feedback and understanding

As discussed in literature reviews and analytical results, there are different types of understanding in spontaneous communication and there are even different levels and scales of understanding within one type (see discussions in Sections 4.1 and 4.2). There are perhaps also different types and levels of micro-feedback (see details in Section 2.10). As Linell (2011) and Clark and Schaefer (1989) have pointed out, the levels of vocal-verbal micro-feedback in relation to understanding can be fuzzy. However, when prosody and gesture are taken into account for the interpretation of the entire micro-feedback unit, which includes both vocal-verbal and gestural components, the relation between micro-feedback and understanding can become clearer. That is, the levels and scales of micro-feedback in relation to understanding are subject to its accompanying gestural and prosodic features as well as to the discourse situation and context.

There is, nevertheless, some kind of understanding involved in some way in all types and levels of micro-feedback expressions. All forms of (comprehensible) micro-feedback imply some kind of understanding, although the understanding can be of very varying depths. Some micro-feedback expressions such as *yeah*, *okay*, head nod, or smile which were labelled as *continued attention* and *acknowledgement* by Clark and Schaefer (1989, p. 267) and *minimal non-verbal and vocal-verbal items*, *single response particles*, and *weak acknowledgement tokens* by Linell (2011, p. 270)

may reflect some kind of shallow (depth of) understanding. Some micro-feedback expressions such as *do you mean this...*, *is it this... but...*, *what did you say*, *sorry*, *I do not understand*, *can you explain*, or the like, which were labelled *demonstration* and *display* by Clark and Schaefer (1989, p. 267), may reflect somewhat increased depths of understanding. Among others, micro-feedback expressions such as *certainly*, *great*, *yes but ...*, *no because ...* which were labelled *stronger acknowledgements*, *repetition of (parts of) prior speaker's utterance*, *(partial) disagreement*, and *stronger disagreements* by Linell (2011, p. 270), may show even deeper levels of understanding. It seems that the weaker and shallower understanding is, the smoother and more even the interaction becomes. This again speaks for the interdependencies between language use, social communicative activity type, interaction situation, participants, and communication goals. For example, when there are more discrepancies in knowledge and experience of participants and more conflicts in communication goals, perhaps more varying levels and scales of understanding would appear in spontaneous communication.

The thesis includes all types of micro-feedback and sometimes other responsive interactions, primarily meaning repair, when they are related to understanding issues. These responsive actions and interactions are involved in understanding to varying extents. Even the same micro-feedback expression can indicate different levels of understanding. For instance, micro-feedback *yeah*, *okay*, and head nod can sometimes signal sufficient understanding of a strong level and sometimes of a weaker level, depending on other factors such as the prosody, the embodied emotions and attitudes, and the discourse context. Even though micro-feedback expressions such as *yeah*, *okay*, and head nod from the main listener might be sometimes seen as primarily “carry-on” signals by the main speaker, the main speaker makes a judgement of these “carry-on” signals and see them primarily as signals for how the information presented has been understood. Then, the main speaker arrives at an understanding of the main listener’s understanding, and subsequently carries on the interaction by either continuing the same topic with further clarifications, specifications, and explanations or developing or changing the interaction to another topic. That is, the so called “carry-on” signal does not simply carry on the conversation but also shows an understanding of the message presented, which contributes to the main speaker’s understanding of the current status of the shared understanding and then helps the main speaker to relate to his or her own intention and make a judgement on how to proceed with the interaction.

All in all, understanding has weaker and stronger levels (even within the same type of understanding classified in this thesis) that are associated with different levels of micro-feedback. Although different levels of micro-feedback and understanding

are not analysed in depth in this thesis, the distinction between acknowledged understanding and demonstrated understanding and even other different depths of understanding should be recognised.

### 12.4.3 Micro-feedback as a means to analyse understanding

This thesis has investigated the relation between micro-feedback and understanding and use of micro-feedback as means for analysing understandings. On the one hand, all micro-feedback expressions communicate some sort of understanding in interaction. On the other, most understanding is communicated through or in relation to micro-feedback. In the empirical data studied, certain vocal-verbal and gestural micro-feedback expressions have been identified as signals for specific types of understanding; also, specific prosodic features of micro-feedback sometimes associated with certain emotions and attitudes have been identified as indicators of specific understanding problems. This interdependency between micro-feedback (i.e., a form) and understanding (i.e., a function and meaning) has been discussed based on the theories of contextualisation and relevance in both Study 1 and Study 2. This is a methodological issue of how to analyse understanding by means of micro-feedback, which is one way among others.

As presented in this thesis, using the knowledge of micro-feedback as a means to study understanding provides a possibility of being able to classify and evaluate understanding in interaction. However, this approach is subject to audio and video analysis and the annotator's and coder's interpretations. Cognitively, it is almost impossible to interpret correctly all the time what is intended and anticipated in terms of sense-making and meaning inferencing in interaction. The thesis primarily focuses on the hearer's understanding of the speaker's information, which is the basis of mutual and shared understanding (if any). This approach can be to some extent constructive for researching the subject of understanding in communication in a more practical way. However, this approach tends to lean towards a position that although understanding is a dynamic and unfinalisable process, it is still linguistically preferred as a static relation to a fact (see also Allwood, 1986, p. 10), which is usually presented in an earlier utterance or discourse context. This approach may have made understanding in interaction countable and accountable for a similar purpose as that of this thesis, although this complex cognitive and linguistic process of understanding in another person's mind may not be so countable and accountable in practice.

### 12.4.4 Language effect

There are some language effects on the participants' communicative behaviours.

## Representativeness of second language users of English

The participants in this thesis project communicated in English. English here is not British, Australian, American, or Canadian. Instead, it is the English lingua franca, which is widely used in an enormous variety of social and cultural contexts (Mauranen & Ranta, 2009). Apparently, as a communication tool, the English language (lingua franca) makes it possible for its second language users from different cultural and language backgrounds to achieve mutually understandable and co-constructive communication. However, it is difficult to make a claim as to how representative the findings can be of such a study that is based on non-native English speakers' use of the English language. It is not easy to measure to what extent the fact that neither of the participants were using their first language in the study affects their communication and understanding. This thesis focuses particularly on the Swedish and the Chinese second language speakers of English. There is a question how generalisable the results of the studies can be when applying them to other peoples in the world.

## Native language influence

It is difficult to speculate which way and to what extent the participants' native language interferes with their use of the English language in intercultural communications. For instance, from a phonological perspective, as known, Chinese is one of the biggest tone languages and Swedish is also a language with tone accents and variations. In addition, both of them have many dialects. Thus, the participant's native language may affect the prosody of his or her spoken English. Similarly, the participant's native language and culture have an impact on how he or she communicates with gestures when speaking in English as a second language. How and how much the native language influences requires further research.

### 12.4.5 Ecological validity discussion: how natural is the research data?

First, this thesis was based on two sets of situated conversations, and the interaction participants were audio- and video-recorded by three cameras in a studio. Thus, the participants were aware of the cameras. It is possible that this situated communication setting had some impact on the participants' communicative behaviours. Given that in Study 1 participants were not previously acquainted and communicated to get to know each other and that in Study 2 participants were previously acquainted with one another and collaborated to solve some learning tasks together, it is not clear how natural or unnatural the conversations and the participants' communicative behaviours actually were. Secondly, the participants were given general instructions about the communication task in both Study 1 and 2. This possibly became an

external motivation for the participants instead of merely something that they themselves wanted. Thus, the data are not perfectly natural in the sense that the conversations were specifically situated and the participants were externally motivated, although the situations were in fact very similar to the naturally occurring ones. Of course, a purely natural empirical body of data of such FTF and also VMC spontaneous conversations, which is good enough for prosody and gesture related understanding analysis, is probably too ideal to aim for.

#### 12.4.6 Other limitations

There are also some other limitations with respect to techniques that are a consequence of the audio- and video-recorded data themselves. Although compared with the secondary data of interviews, questionnaires, and audio (only) recordings in this thesis, audio- and video-recordings as the primary data to some extent may capture more information about what is really going on with the participants, there might still be something missing or not possible to capture. For example, it is not uncommon that some minimal gestures and spoken words cannot clearly be perceived and transcribed, often because of too extensive or intensive gestural movements and too many interfering interactional overlaps and echoes. Besides these, the coding schemes that were used in this thesis might be restricted to what is studied and what can be consequently found, and the transcriptions and annotations were based on the transcribers and annotators' interpretations.

### 12.5 Suggestions for future studies

A few suggestions can be made for future studies. First, in this thesis, micro-feedback has been primarily investigated in terms of its modality and prosody in relation to understanding, thus the gestural aspects of micro-feedback of relevance could be studied further in the future. As discussed in the earlier chapters, many theories of pragmatics and a number of language and communication researchers take the vocal-verbal content of utterances as the basic unit of communication and have neglected, relatively speaking, the fact that the gestural content also provides very important information about the meaning of the utterance. As Navarretta and Paggio (2013) stressed, the physical setting, the number of participants, the topic discussed, and the degree of *familiarity* (Campbell, 2007) influence the use of micro-feedback and particularly its related gestures. Thus, gestural micro-feedback can be studied in such a larger variety of contexts, as mentioned above, for studies of both communication and understanding. Also, multimodal behaviour of micro-feedback associated with understanding and understanding problems, for example, how gestural micro-

feedback complements vocal-verbal micro-feedback in expressing different understandings, could be investigated in various communication contexts.

Secondly, studying how gestures and prosody relate to the same micro-feedback expression in a freer communication setting could be of interest in the future. As Prieto, Borrás-Comes, Tubau, and Espinal (2013) claimed, there is a direct association between gestural and prosodic features linked with specific linguistic phenomena (e.g., micro-feedback). Both prosody and gesture constrain meanings in guiding the speakers to interpret what has been understood or not. As Prieto et al. (2013) noted, gestural and prosodic features are not only important as a helping hand in interpreting multimodal communication and understanding, they also interact with each other during the interpreting process. Therefore, how gesture and prosody are related to understanding through micro-feedback could be of interest for investigation in the future.

Thirdly, this thesis has investigated micro-feedback in relation to understanding. How micro-feedback and understanding are related to turn management is another interesting question. Schegloff (1998) pointed out that prosodic features of speech had certain associations with communication functions, for example, turn management. In Allwood's (2000) theory of interactive communication management (ICM), management of turns and *feedback* (i.e., *micro-feedback*) are two important ICM functions of interaction. Kern (2007) also claimed that turn constructional units with a final falling-to-mid pitch in general terminate the information conveyed and project or encourage responses that signal understanding more explicitly. In addition, turn constructional units with a final high-pitch do not only indicate that the turn will be continued, they also create sequential positions for minimal responses such as micro-feedback (Kern, 2007). Additionally, Navarretta and Paggio (2013) also found that all types of head movement were used as *feedback* and many of them were involved in turn management, which suggests that *feedback* expressions may simultaneously have the function of turn management. Although here, what is a turn and how it interacts with micro-feedback and understanding in the communication construction process will not be conceptualised, it seems that there is some kind of relation or interaction between them and this is little explained in conversation analysis theories. Studying micro-feedback and understanding in relation to turn management could be of interest in the future.

Moreover, as presented within the framework for analysing the concept of micro-feedback, emotional and attitudinal reactions are also important communicative functions of micro-feedback that help to evaluate understandings. In addition, prosodic cues are important for emotional and attitudinal interpretations in social conversation (Couper-Kuhlen, 2009). Thus, the emotional and attitudinal functions of

micro-feedback and its associated prosody as well as other relevant aspects in relation to understanding could be further explored in future studies.

Next, concerning the research method, as Couper-Kuhlen and Selting (1996) and House (2006) claimed, a pragmatic account of prosody in understanding must ultimately be based on and supported by empirical data from natural interactions. More natural conversations in a more natural situation or setting than has been studied in this thesis will have to be taken into account in future research. In order to have a natural conversation while still capturing the audio and video information, more commonly used (e.g., day-to-day) audio and video information technologies such as mobile phones and tablets could be used. This would result in more natural audio and video conversational data, which will allow for more detailed and empirical analyses of understanding in interaction.

In addition, the position of the micro-feedback unit in the utterance (e.g., beginning, middle, or end), the pitch changing tendency of the micro-feedback (e.g., increasing, decreasing, or sustaining), and the relevant topics based on more (than three) categorisations of pitch range type and duration type that have been studied in this thesis could also be of interest.

Besides these, deeper scales of micro-feedback and understandings could be of interest to research. Further categorisations of micro-feedback and understanding than those presented in this thesis could be investigated in greater detail.

## **12.6 Concluding remarks**

The overarching aim of this thesis is to contribute to understanding the understanding in real-time communication by empirically investigating how understanding is signalled, detected, handled, and resolved in social interactions of varying complexity in intercultural, multimodal, and video-mediated communication situations. The analytical focuses are on micro-feedback and meaning repair, using an interactional approach based on theories of social communicative activity type, meaning and implicature, contextualisation, and relevance. The thesis also aims to uncover similarities and differences in understanding between face-to-face and video-mediated communication.

This thesis comprises two major empirical studies. Study 1 aims to investigate micro-feedback in relation to understanding issues in a spontaneous communication activity in first encounters. Based on the results from Study 1, Study 2 expands the research and aims to examine how understanding problems are coped with by acquainted interlocutors in relation to not only micro-feedback but also meaning repair in an educational activity with collaborative learning tasks.

The empirical data consist of intercultural, multimodal, and technology-mediated communication, where the English lingua franca is spoken, between Swedish and Chinese speakers. This is because the Swedish and the Chinese have significant physical, regional, linguistic, and socio-cultural differences, which according to the earlier theories (e.g., Gumperz, 1982; Tannen, 1990; Samovar et al., 2012) have more individual differences in communication presuppositions, expectations, and common knowledge and resources in sense-making (Linell, 2009) and thus likely find it more difficult to achieve understanding. Besides this, contextual and technological influences on communication and understanding have been recognised in different communicative activities. Understanding is more salient in communication of complex tasks than simple ones (e.g., Lindwall et al., 2005; Sins et al., 2011). Equally important, some research has found that compared to FTF, VMC is more constrained when it comes to achieving understanding (e.g., Olson & Olson, 2000), while there is also research that has found that mediating technology has little effect on interaction and understanding (e.g., Anderson, 2006). The present study attempts to investigate understanding in real-time communication with specific focuses on all these elements.

This thesis gives an account of how to evaluate understanding through one of its signals—micro-feedback—and the related meaning repair interactions and how to operationalise understanding through micro-feedback and meaning repair in social interaction. The study is not limited to presenting the properties of micro-feedback and understanding. It also affords or facilitates the production and intelligibility of the accounts of micro-feedback and understanding. As Lynch (2011, p. 555) pointed out, “the production and display of understandable materials” (such as micro-feedback in particular in this thesis) “provide reciprocal conditions for understanding” whether anything is understood and how (much) it is understood, regardless of whether it is “accomplished by participants or analysts”. Viewed from Lynch’s perspective, a good study of this issue is more valuable if it illuminates the interactional production of the material conditions and exhibitions (e.g., micro-feedback modality and prosody and the related meaning repair interactions) for understanding rather than only outlining what it is.

Besides the specific patterns and features of how micro-feedback is related to different types of understandings, how understanding is classified, and how understanding problems are detected, handled, and resolved as answers to the thesis research questions, this thesis has also found that misunderstandings do not occur as frequently as they are predicted in intercultural communications, either in the first encounters or in the task-solving collaboration. Visual modality (gesture) plays an

important role in spontaneous communication, in particular in the head region focusing on head movements and facial expressions. Unimodal head movement micro-feedback exclusively signals sufficient understanding. Gestural micro-feedback of eyebrow rise or frown, head forward, gaze sideways (from) or gaze at (the other interlocutor), and sometimes smile as well as multimodal micro-feedback chuckle and laughter can be indicators of understanding problems. *Yeah* and nod are typically signals of sufficient understanding but are sometimes related to misunderstanding, which is often associated with hesitation and uncertainty in the prosody. Based on the empirical data, when information is repeated, paraphrased, or responded to with unanticipated actions, a misunderstanding may have occurred. Misunderstanding can, but does not necessarily, lead to sufficient understanding. As a matter of fact, it can cause more misunderstandings in the interaction even without the interlocutors' awareness. The thesis has also found that video mediating technology does not seem to affect understanding or understanding problems, although it is easier to detect and repair understanding problems in FTF than in VMC. People have higher interdependency and interactivity in FTF than in VMC, and FTF provides better chances of detecting, handling, and resolving understanding problems than does VMC.

Micro-feedback, as this thesis clearly suggests, does not always signal or determine understanding in itself. The relation between micro-feedback and understanding is not that simple, it is complex and multifaceted, with many interrelated and possibly also overlapping components perhaps both known (e.g., modality and prosody, and sometimes meaning repair) and unknown (e.g., turn management). What can be claimed, and also what has been demonstrated in the thesis, is that the relation between micro-feedback and understanding is highly dependent on the contextualisation of relevance for sense-making and meaning inferencing. On the one hand, contextualisation assists in understanding the phenomenon of micro-feedback and the issue of understanding in communication. On the other hand, micro-feedback and its prosodic and gestural features contribute to the process of contextualisation. As understanding is an interactive and situated phenomenon shown by the use of language and the construction of discourse, the social communicative activity type (Wittgenstein's (1958) *language games*, Allwood's (1976) *behaviour, situation, and meaning types*, Levinson's (1979) *activity type*, Allwood's (2013) *activity-based communication analysis*, and Linell's (2010) *communicative activity type*), the implication of intended meaning and anticipated reaction (Grice's (1975) *meaning and implicature*), the context of interaction (Gumperz' (1982) *contextualisation*), and the discourse of relevance (Sperber and Wilson's (1986) *relevance theory*) should all be

taken into account when studying understanding in communication. That is, context, discourse, micro-feedback (both modality and prosody), meaning repair, and understanding are interdependent in social interaction. Micro-feedback has an undeniable role in relation to understanding and it also contributes to the management of communication.

The findings and discussions in this thesis constitute a new starting point for research on understanding in real-time communication through micro-feedback and meaning repair. Further research could possibly strengthen and extend the findings beyond the limitations of this study in culture, language, communication activity, and technology.

Apart from enhancing the theoretical understanding of understanding in real-time communication, the empirical findings in this thesis can contribute a practical basis for developing guidelines and strategies for the design of communication technology applications, such as systems for speech, gesture and understanding recognition, graphical display, and motion capture and simulation. The results can also add to the foundation for practical design of technology enhanced education and communication, for example, online and flexible learning and digital communication, not least in intercultural settings.

# References

- Abercrombie, D. (1965). *Studies in phonetics and linguistics*. London: Oxford University Press.
- Abrams, Z. I. (2008). Sociopragmatic features of learner-to-learner computer-mediated communication. *CALICO journal*, 26(1), 1–27. doi:10.11139/cj.28.2.369-391
- AbuSeileek, A., & Qatawneh, K. (2013). Effects of synchronous and asynchronous computer-mediated communication (CMC) oral conversations on English language learners' discourse functions. *Computers & Education*, 62, 181–190. doi:10.1016/j.compedu.2012.10.013
- Allenmark, F., Hsu, Y.-F., Roussel, C., & Waszak, F. (2015). Repetition priming results in sensitivity attenuation. *Brain Research*, 1626, 211–217.
- Allport, F. (1924). *Social psychology*. Boston: Houghton-Mifflin.
- Allwood, J. (1976). *Linguistic communication as action and cooperation*. Gothenburg Monographs in Linguistics 2. University of Gothenburg, Department of Linguistics.
- Allwood, J. (1986). Some perspectives on understanding in spoken interaction. In M. Furberg, T. Wetterström, & C. Åberg (Eds.), *Logic and abstraction* (pp. 1–30). Acta Philosophica Gothoburgensia 1.
- Allwood, J. (1993). Feedback in second language acquisition. In C. Perdue (Ed.), *Adult language acquisition: Cross linguistic perspectives, vol. II* (pp. 196–235). Cambridge: Cambridge University Press.
- Allwood, J. (2000). An activity based approach to pragmatics. In H. Bunt, & W. Black (Eds.), *Abduction, belief and context in dialogue: Studies in computational pragmatics* (pp. 47–80). Amsterdam: John Benjamins.
- Allwood, J. (2002). Bodily communication: Dimensions of expression and content. In B. Granström, D. House, & I. Karlsson (Eds.), *Multimodality in language and speech systems* (pp. 7–26). Dordrecht: Kluwer Academic Publishers.
- Allwood, J. (2013). A multidimensional activity based approach to communication. In I. Wachsmuth, J. de Ruiter, P. Jaecks, & S. Kopp. (Eds.) *Alignment in communication (Vol. 6)* (pp. 33–55). Amsterdam: John Benjamins.
- Allwood, J. (2015). English translation of Tvärkulturell kommunikation (1985), *Papers in Anthropological Linguistics 12*, University of Göteborg, Danish Intercultural Organization, 17/04/2015.
- Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., & Paggio, P. (2007). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources & Evaluation*, 41(3–4), 273–287. doi:10.1007/s10579-007-9061-5
- Allwood, J., & Lu, J. (2011). Unimodal and multimodal co-activation in first encounters: A case study. In P. Paggio, E. Ahlsén, J. Allwood, K. Jokinen, & C. Navarretta (Eds.), *Proceedings of the 3rd Nordic Symposium on Multimodal Communication* (pp. 1–9). University of Helsinki, Finland, 27–28 May 2011. NEALT Northern European Association for Language Technology Proceedings Series, Vol. 15.
- Anderson, A. H. (2006). Achieving understanding in face-to-face and video-mediated multiparty interactions. *Discourse Processes*, 41(3), 251–287. doi:10.1207/s15326950dp4103\_2
- Anward, J., & Linell, P. (2016). On the grammar of utterances: putting the form vs. substance distinction back on its feet. *Acta Linguistica Hafniensia*, pp. 1–24.
- Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 555–596.
- Austin, J. L. (1962). *How to do things with words* (2nd Ed.). Urmson J.O., & Sbisá. M. (Eds.). Cambridge: Harvard University Press.
- Bakhtin, M. M. (1984). *Problems of Dostoevsky's Poetics*. Edited by Emerson, C. Minneapolis, MN: University of Minnesota Press.
- Bakhtin, M. M. (1986). *Speech genres and other late essays*. C. Emerson, & M. Holquist (Eds.), V. W. McGee (Trans.). Austin: University of Texas Press.
- Bakhtin, M. M. (1992). *The dialogic imagination: Four essays*. Austin: University of Texas Press.

- Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. S. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 87, 156–179.
- Barron, A. (2003). *Acquisition in interlanguage pragmatics: Learning how to do things with words in a study abroad context*. Philadelphia & Amsterdam: John Benjamins.
- Bateson, G. (1956). The message “this is play”. In B. Schaffner (Ed.), *Group processes* (pp. 145–242). New York: Josiah Macy Jr. Foundation.
- Bauman, R., & Briggs, C. (1990). Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology*, 19, 59–88.
- Bazzanella, C., & Damiano, R. (1999). The interactional handling of misunderstanding in everyday conversations. *Journal of Pragmatics*, 21(6), 817–836.
- Bertrand, R., & Goujon, A. (2017). (Dis)aligning for improving mutual understanding in talk-in-interaction. *Revue française de linguistique appliquée*, xxii(2), 53–70.
- Blöndal, Þ. (2005). Feedback in conversational storytelling. In J. Allwood (Ed.), *Feedback in spoken interaction – Nordtalk Symposium 2003. Gothenburg Papers in Theoretical Linguistics*, 91 (pp. 1–17). Göteborg University, Dept. of Linguistics.
- Blum-Kulka, S., & Weizman, E. (1988). The inevitability of misunderstandings: Discourse ambiguities. *Text*, 8(3), 219–241.
- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (version 5.1.05). Retrieved May 1, 2009, from <http://www.praat.org/>
- BonJour, L. (2002). *Epistemology: Classic problems and contemporary responses*. Lanham, MD: Rowman & Littlefield.
- Borod, J. C. (1993). Cerebral mechanisms underlying facial, prosodic, and lexical emotional expression: A review of neuropsychological studies and methodological issues. *Neuropsychology*, 7, 445–463.
- Boud, D. J., & Molloy, E. (Eds.). (2013). *Feedback in higher and professional education: Understanding it and doing it well*. New York: Routledge.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41, 687–699.
- Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge, England: Cambridge University Press.
- Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.
- Campbell, N. (2007). Individual traits of speaking style and speech rhythm in a spoken discourse. In *Proceedings of the COST 2012 Workshop* (pp. 107–120).
- Cerrato, L. (2005). On the acoustic, prosodic and gestural characteristics of “m-like” sounds in Swedish. In J. Allwood (Ed.), *Feedback in Spoken Interaction – Nordtalk Symposium 2003. Gothenburg Papers in Theoretical Linguistics*, 91 (pp. 18–31). Göteborg University, Dept. of Linguistics.
- Cerrato, L., & Skhiri, M. (2003). Analysis and measurement of communicative gestures in human dialogues. In *Proceedings of AVSP 2003* (pp. 251–256). France.
- Chomsky, N. (1995). *The minimalist program*. (Current studies in linguistics 28.) Cambridge, MA: MIT Press.
- Clark, H. H. (1993). *Arenas of language use*. Stanford: The University of Chicago Press.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, D. C.: American Psychological Association.
- Clark, H. H., & Schaefer, E. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1–39.
- Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Connerley, M. L., & Pedersen, P. B. (2005). Chapter 3: Cultural frameworks and their importance for leaders. *Leadership in a Diverse and Multicultural Environment: Developing Awareness, Knowledge, and Skills*. CA: SAGE Publications. doi:10.4135/9781483328966.n3
- Couper-Kuhlen, E. (1996). The prosody of repetition: On quoting and mimicry. In E. Couper-Kuhlen, & M. Selting (Eds.), *Prosody in conversation* (pp. 366–405). Cambridge: Cambridge University Press.

- Couper-Kuhlen, E. (2001). Intonation and discourse: Current reviews from within. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 13–34). New York and Oxford: Blackwell Publishers.
- Couper-Kuhlen, E. (2009). A sequential approach to affect: The case of 'disappointment'. In M. Haakana, M., Laakso, & J., Lindström (Eds.), *Talk in interaction-comparative dimensions* (pp. 94–123). Helsinki: Finnish Literature Society (SKS).
- Couper-Kuhlen, E., & Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In E. Couper-Kuhlen, & M. Selting (Eds.), *Prosody in conversation*. (pp. 11–56). Cambridge: Cambridge University Press.
- Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge, UK: Cambridge University Press.
- Cutrone, P. (2010). The backchannel norms of native English speakers: A target for Japanese L2 English learners. *Language Studies Working Papers 2*, 28–37.
- Dannerer, M. (2004). Misunderstanding at work. In K. Aijmer (Ed.), *Dialogue Analysis VIII: Understanding and Misunderstanding in Dialogue: Selected Papers from the 8th IADA Conference* (pp. 103–118), Göteborg 2001. Druck & Einband: Walter de Gruyter.
- Dascal, M. (1985). The relevance of misunderstanding. In M. Dascal (Ed.), *Dialogue: An interdisciplinary approach* (pp. 441–459). Amsterdam: Benjamins.
- Dascal, M. (1999). Introduction: Some questions about misunderstanding. *Journal of Pragmatics*, 31(6), 753–762.
- Dascal, M. (2003). *Interpretation and understanding*. Philadelphia & Amsterdam: John Benjamins Publishing.
- Dascal, M., & Berenstein, I. (1987). Two modes of understanding: Comprehending and grasping. *Language and Communication*, 7, 139–151.
- de Mooij, M. (2013). *Human and mediated communication around the world: a comprehensive review and analysis*. New York & London: Springer Science & Business Media. doi:10.1007/978-3-319-01249-0
- Deng, X. D. (2009). Listener response. In S. D'Hondt, J. Östman, & J. Verschueren (Eds.), *The pragmatics of interaction* (pp. 104–124). Amsterdam: John Benjamins Publishing Company.
- de Souza, F., & Danilo, M. (1985). Dialogue breakdowns. In M. Dascal (Ed.), *Dialogue: An interdisciplinary approach* (pp. 415–426). Amsterdam and Philadelphia: John Benjamins.
- Dodd, S. (2005). *Understanding autism*. Marrickville: Elsevier Australia.
- Draghici, A. (2007). *Human resources management*, Editura Politehnica, Timisoara.
- Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors*, 45, 445–454. doi:10.1518/hfes.45.3.445.27258
- Duncan, S., & Fiske, D. (1977). *Face-to-face interaction: Research, methods, and theory*. Hillsdale, NJ. Lawrence Erlbaum.
- Durkheim, E. (1895). Les règles de la méthode sociologique, S. Soloway and J. Mueller (Trans.), G. Cattin (Ed.), as Durheim, E (1938) *The rules of sociological method*. Glencoe, Illinois: The Free Press.
- Edwards, P. (Ed.) (1967). *Encyclopedia of philosophy*. New York, NY: Macmillan.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6, 169–200.
- Eliot, T. S. (2010). *Notes towards the definition of culture*. London: Faber & Faber Publisher.
- Ermakova, O. P., & Zemskaja, E. A. (1993). K postroeniiu tipologii kommunikativnykh neu-dach (na materiale estestvennogo russkogo iazyka) [On constructing a typology of communicative failures on the basis of authentic Russian material]. In *Russkii iazyk i ego funktsionirovanie: kommunikativno-pragmaticheskii aspekt* (pp. 90–157). Moscow: Nauka.
- Fägersten, K. B., Holmsten, E., & Cunningham, U. (2010). Multimodal Communication and Meta-modal Discourse. In R. Taiwo (Ed.), *Handbook of Research on Discourse Behavior and Digital Communication: Language Structures and Social Interaction: Language Structures and Social Interaction* (pp. 146–163). Hershey & New York: Information Science Reference IGI Global.
- Farr, R. (1996). *The roots of modern social psychology 1872–1954*. Oxford: Blackwell.
- Finkbeiner, R., Meibauer, J., & Schumacher, P. B. (Eds.). (2012). *What is a context?: Linguistic approaches and challenges*. Philadelphia, PA: John Benjamins.
- Firth, J. R. (1957). *Papers in linguistics 1934–1951*. London: Oxford University Press.
- Fisher, D., & Frey, N. (2007). *Checking for understanding: Formative assessment techniques for your classroom*. Alexandria, VA: ASCD Publishing.
- Fishman, P. (1978). Interaction: The work women do. *Social Problems*, 25, 397–406.

- Fleiss, J. L. (1971) Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382.
- Fraser, B. (1993). No conversation without misrepresentation. In H. Parret (Ed.), *Pretending to communicate* (pp. 143–153). Berlin: Walter de Gruyter Verlag.
- Friedman, R. A., & Currall, S. C. (2003). E-mail escalation: Dispute exacerbating elements of electronic communication. *Human Relations*, 56(11), 1325–1347. doi:10.2139/ssrn.304966
- Garfinkel, H. (1967). *Studies in ethnomethodology*. (Prentice-Hall). New ed. 1984 [1992]. Cambridge: Polity Press.
- Garner, J. T., & Buckner, M. M. (2013). Skyping class: Using videoconferencing in organizational communication classes. *Communication Teacher*, 27(1), 1–5. doi:10.1080/17404622.2012.730618
- Goffman, E. (1971). *Relations in public*. Allen Lane, Harmondsworth: The Penguin Press.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. New York, NY: Harper & Row.
- Gogan, L. M., Popescu, A. D., & Duran, V. (2014). Misunderstanding between cross-cultural members within collaborative engineering teams. *Procedia-Social and Behavioral Sciences*, 109, 370–374.
- Goodwin, C. (1981). *Conversational organization: Interactions between speakers and hearers*. New York: Academic Press.
- Goodwin, C. (2018). *Co-operative action. Learning in doing: Social, cognitive and computational perspectives*. NY: Cambridge University Press.
- Grammer, K., Kopp, S., Allwood, J., & Ahlsén, E. (2008). A framework for analyzing embodied communicative feedback. In J. Martin (Ed.), *Multimodal Corpora. JLRE (Special Issue on Multimodal Corpora)* (pp. 7–22).
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics, vol. 3, speech acts* (pp. 41–58). New York: Academic Press.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge: Harvard University Press.
- Gulick, W. L., Gescheider, G. A., & Frisina, R. D. (1989). *Hearing: Physiological acoustics, neural coding, and psychoacoustics*. New York: Oxford University Press.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
- Hacking, I. (2004). *Historical ontology*. Cambridge: Harvard University Press.
- Hall, E. T. (1959). *The silent language*. NY: Doubleday.
- Hancock, J., & Dunham, P. (2001). Language use in computer-mediated communication: The role of coordination devices. *Discourse Processes*, 31, 91–110.
- Hård af Segerstad, Y. (2002). *Use and adaptation of written language to the conditions of computer-mediated communication*. Doctoral dissertation, University of Gothenburg.
- Hartnack, J. (2013). *Wittgenstein and modern philosophy*. London & New York: Routledge, Taylor and Francis Group.
- Hatem, W. A., Kwan, A., & Miles, J. (2012). Comparing the effectiveness of face to face and computer mediated collaboration. *Advanced Engineering Informatics*, 26, 383–395. doi:10.1016/j.aei.2012.01.001
- Hayakawa, H. (2003). *The meaningless laugh: Laughter in Japanese communication*. Unpublished thesis of Doctor of Philosophy at University of Sydney.
- Helmstad, G. (1999). *Understanding s of understanding an inquiry concerning experiential conditions for developmental learning*. Acta universitatis gothoburgensis.
- Hindmarsh, J., Reynolds, P., & Dunne, S. (2011). Exhibiting understanding: The body in apprenticeship. *Journal of Pragmatics*, 43(2), 489–503.
- Hirst, D. J. (1999). The symbolic coding of duration and alignment: An extension to the INTSINT system. In *Proceedings of Eurospeech '99*. Budapest, September.
- Hirst, D. J., & Di Cristo, A. (Eds.). (1998). *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.
- Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. (1994). Repairing conversational misunderstandings and non-understandings. *Speech Communication*, 15, 213–229.
- Holliday, A., Hyde, M., & Kullman, J. (Eds.). (2004). *Intercultural communication: An advanced resource book*. London: Routledge.
- Holliday, A., Kullman, J., & Hyde, M. (Eds.). (2016). *Intercultural communication: An advanced resource book for students (Routledge Applied Linguistics)* (3rd Ed.). New York: Taylor & Francis.
- Homans, G. L. (1961). *Social behavior: its elementary forms*. London: Routledge, and Kegan Paul.
- Horton, W. S., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Processes*, 40, 1–35.

- House, J. (2006). Constructing a context with intonation. *Journal of Pragmatics*, 38, 1542–1558.
- Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont, CA: Thomson Wadsworth.
- Hutchby, I. (2001). *Conversation and technology: From the telephone to the internet*. Cambridge: Polity Press.
- Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis: Principles, practices, and applications*. Cambridge: Polity Press.
- Ishi, C. T., Ishiguro, H., & Hagita, N. (2014). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, 57, 233–243.
- Jurafsky, D., Shriberg, E., Fox, B., & Curl, T. (1998). Lexical, prosodic, and syntactic cues for dialog acts. In *Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers* (pp. 114–120).
- Kaur, J. (2010). Achieving mutual understanding in world Englishes. *World Englishes*, 29(2), 192–208.
- Kempson, R. M. (1977). *Semantic theory: Cambridge textbook in linguistics*. Cambridge & New York: Cambridge University Press.
- Kennedy, M. R. T., & DeRuyter, F. (1991). Cognitive and language bases for communication disorders. In D. R. Beukelman, & K. M. Yorkston (Eds.), *Communication disorders following traumatic brain injury: Management of cognitive, language, and motor impairments* (pp. 123–190). Austin, TX: Pro-Ed.
- Kern, F. (2007). Prosody as a resource in children's game explanations: Some aspects of turn construction and reciprocity. *Journal of Pragmatics*, 39, 111–133.
- Kiesler, S., & Sproull, L. (1992). Group decision making and communication technology. *Organizational Behavior & Human Decision Processes*, 52, 96–123. doi:10.1016/0749-5978(92)90047-B
- Kopp, S., Grammer, K., Allwood, J., Ahlsén, E., Oberzaucher, E., & Koppensteiner, M. (2007). The analysis of embodied communicative feedback in multimodal corpora: A prerequisite for behavior simulation. *Journal of Language Resources and Evaluation*, 41, 255–272.
- Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Chichester: Wiley-Blackwell.
- Krippendorff, K. (1993). The past of communication's hoped-for future. *Journal of Communication*, 43(3), 34–44.
- Kushida, S. (2011). Confirming understanding and acknowledging assistance: Managing trouble responsibility in response to understanding check in Japanese talk-in-interaction. *Journal of Pragmatics*, 43(11), 2716–2739.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Leontiev, A. A. (1981). *Psychology and the language learning process*. Oxford: Pergamon.
- Lévinas, E. (1961). *Totality and infinity; an essay on exteriority*. Lingis, A. (Trans.). Pittsburgh: Duquesne University Press.
- Levinson, S. C. (1979). Activity types and language. *Linguistics*, 17, 365–399.
- Lindström, B., Marton, F., Emanuelsson, J., Lindahl, M., & Packendorff, M. (2011). Enhancing arithmetic skills through perceptual - bodily interaction with a computer game. In J. Emanuelsson, L. Fainsilber, J. Häggström, A. Kullberg, B. Lindström, & M. Löwing. (Eds.), *Voices on learning and instruction in mathematics* (pp. 119–142). Göteborg, Nationellt Centrum för Matematik.
- Lindström, N. B. (2008). *Intercultural communication in health care: Non-Swedish physicians in Sweden*. PhD dissertation. Gothenburg Monographs in Linguistics 36.
- Lindwall, O. (2008). *Lab work in science education: Instruction, inscription and practical achievement of understanding*. Linköping studies in arts and science 456. The Department of Theme Research. Linköping, Sweden.
- Lindwall, O., & Lymer, G. (2011). Uses of “understand” in science education. *Journal of Pragmatics*, 43(2), 452–474.
- Lindwall, O., Lymer, G., Lindström, B., & Bernhard, J. (2005). Approaching interactive engagement tasks: grounds for relevance and irrelevance in a computer-based mechanics lab. COSIL, Department of Education, Learning and Teaching Unit.
- Linell, P. (1998). *Approaching dialogue: Talk, interaction and contexts in dialogical perspectives*. Amsterdam & Philadelphia: John Benjamins Publishing.
- Linell, P. (2009). *Rethinking language, mind and world dialogically: Interactional and contextual theories of human sense-making*. Charlotte, NC: Information Age Publishing.
- Linell, P. (2010). Communicative activity types as organisations in discourses and discourses in organisations. In S. K. Tanskanen, M. L. Helasvuo, M. Johansson, & M. Raitaniemi (Eds.), *Discourses in interaction* (pp. 33–59).

- Linell, P. (2011). *Samtalskulturer: Kommunikativa verksamhetstyper i samhället*. Studies in Language and Culture, 18. Linköping: Department of Culture and Communication.
- Linell, P. (2013). The dynamics of incrementation in utterance-building: Processes and resources. In B. Szczepek Reed, & G. Raymond (Eds.), *Units of talk: Units of actions* (pp. 57–89). Amsterdam: John Benjamins.
- Linell, P. (2017). Dialogue, dialogicality and interactivity: A conceptually bewildering field? *Language and Dialogue*, 7(3), 301–335.
- Linell, P. (2018). The dynamics of contexts of sense-making. In D. Favareau, (Ed.), *Co-operative engagements of intertwined semiosis: Essays in honour of Charles Goodwin* (pp. 192–203). Tartu: University of Tartu Press.
- Linell, P. & Persson Thunqvist, D. (2003). Moving in and out of framings: Activity contexts in talks with young unemployed people within a training project. *Journal of Pragmatics*, 35, 409–434.
- Littlejohn, S. W., Foss, K. A., & Oetzel, J. G. (2016). *Theories of human communication* (11th ed.). Illinois: Waveland Press.
- Lu, J. (2010). *A pilot study of Chinese gestural feedback* (Master thesis, University of Gothenburg, Sweden). Retrieved from <http://www.essays.se/essay/8bc518aa6e/>
- Lu, J. (2012). Eliciting gestural feedback in Chinese and Swedish informal interactions. In Y. Nakano, M. Neff, A. Paiva, & M. Walker (Eds.), *Proceedings of Intelligent Virtual Agents 2012* (pp. 480–482). Springer-Verlag.
- Lu, J. (2014). *Modality and prosody of micro-feedback in relation to understanding in intercultural communication-first encounters between Swedes and Chinese*. Licentiate thesis. Gothenburg: Chalmers Repro.
- Lu, J., & Allwood, J. (2011). Unimodal and multimodal feedback in Chinese and Swedish mono-cultural and intercultural interactions (a case study). In P. Paggio, E. Ahlsén, J. Allwood, K. Jokinen, & C. Navarretta (Eds.), *Proceedings of the 3rd Nordic Symposium on Multimodal Communication* (pp. 40–47). Finland, Helsinki, 27–28 May 2011. NEALT Proceedings Series, Vol. 15.
- Lu, S. E. (2017). Coordination-free equilibria in cheap talk games. *Journal of Economic Theory*, 168, 177–208.
- Luckmann, T. (1990). Social communication, dialog, and conversation. In I. Markova, & K. Foppa (Eds.), *The dynamics of dialog* (pp. 46–61). New York: Harvester Wheatsheaf.
- Lustig, M. W., & J. Koester. (2003). *Intercultural competence: Interpersonal communication across cultures*. Boston: Allyn and Bacon.
- Lynch, M. (2011). Commentary: On understanding understanding. *Journal of Pragmatics*, 43(2), 553–555.
- Lyon, C., Nehaniv, C. L., & Cangelosi, A. (Eds.). (2007). *Emergence of communication and language*. London: Springer-Verlag London Limited.
- Macbeth, D. (1994). Classroom encounters with the unspeakable: “Do you see, Danelle?”. *Discourse Processes*, 17, 311–335.
- Mahboob, A., & Knight, N. (Eds.). (2010). *Applicable linguistics*. London & New York: Continuum International Publishing Group.
- Marková, I. (2003). *Dialogicality and social representations*. Cambridge: Cambridge University Press.
- Martínez-Flor, A., & Usó-Juan, E. (Eds.). (2010). *Speech act performance: Theoretical, empirical and methodological issues*. Philadelphia & Amsterdam: John Benjamins.
- Mauranen, A., & Ranta, E. (Eds.). (2009). *English as a lingua franca: Studies and findings*. Newcastle upon Tyne, NE, UK: Cambridge Scholars Publishing.
- Maynard, D., & Zimmerman, D. H. (1984). Topical talk, rituals, and the social organization of relationships. *Social Psychology Quarterly*, 47(4), 301–316.
- McConnell, C. R. (1993). *The health care supervisor on effective communication*. Gaithersburg, Maryland: Aspen Publishers.
- McHale, J. P. (2004). *Communicating for change: Strategies of social and political advocates*. Oxford: Rowman & Littlefield Publishers.
- McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo, & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9–45). San Francisco: Jossey-Bass. doi:10.1108/17410400410533908
- Mead, G. H. (1934). *Mind, self and society*. Chicago: The University of Chicago Press.
- Mead, G. H. (1982). *The Individual and the social self: Unpublished essays by G. H. Mead*. D. L. Miller (Ed.). Chicago: University of Chicago Press.

- Merola, G., & Poggi, I. (2004). Multimodality and gestures in the teacher's communication. In A. Camurri, & G. Volpe (Eds.), *Gesture-based communication in human-computer interaction: Lecture notes in computer science*, vol. 2915 (pp. 101–111). Springer Berlin Heidelberg.
- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (Eds.). (2005). *Assessing science understanding: A human constructivist view*. Elsevier Academic Press.
- Mitchell, R. L. C., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience and Biobehavioral Reviews*, 37, 471–479.
- Moeschler, J. (1985). *Argumentation et conversation*. Paris: Cr dific-Hatier.
- Mondada, L. (2011). Understanding as an embodied, situated and sequential achievement in interaction. *Journal of Pragmatics*, 43(2), 542–552.
- Monk, A. (2009). *Common ground in electronically mediated communication*. Morgan & Claypool Publishers.
- Mustajoki, A. (2012). A speaker-oriented multidimensional approach to risks and causes of miscommunication. *Language and Dialogue*, 2(2), 216–243.
- Nadeu, M., & Prieto, P. (2011). Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics*, 43, 841–854.
- Naess, A. (1966). *Communication and argument*. London: Allen & Unwin.
- Navarretta, C., Ahls n, E., Allwood, J., Jokinen, K., & Paggio, P. (2012). Feedback in Nordic first-encounters: A comparative study. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. European language resources distribution agency (pp. 2494–2499).
- Navarretta, C., & Lis, M. (2012). Multimodal feedback expressions in Danish and Polish spontaneous conversations. In *The Northern European Association for Language and Technology NEALT Proceedings of the fourth Nordic Symposium of Multimodal Communication* (pp. 55–62), November 2012, Gothenburg Sweden. doi:10.1007/978-3-319-07230-2\_44
- Navarretta, C., & Paggio, P. (2010). Classification of feedback expressions in multimodal data. *Proceedings of the ACL 2010 Conference Short Papers* (pp. 318–324). Uppsala, Sweden, 11–16 July 2010. Association for Computational Linguistics.
- Navarretta, C., & Paggio, P. (2013). Multimodal turn management in Danish dyadic first encounters. In *Proceedings 4th Nordic Symposium on Multimodal Communication, Northern European Association for Language and Technology*, (pp. 5–12), November 15–16, Gothenburg, Sweden.
- Navarretta, C., Paggio, P., & Jokinen, K. (2008). Distinguishing the communicative functions of gestures. In A. Popescu-Belis, & R. Stiefelhagen (Eds.), *Proceedings of 5th Joint Workshop on Machine Learning and Multimodal Interaction* (pp. 38–49), Utrecht, Springer.
- Neuliep, J. W. (2016). *Intercultural communication: A contextual approach* (7th ed.). NY & London: SAGE Publications.
- Newlands, A., Anderson, A., & Mullin, J. (2003). Adapting communicative strategies to computer-mediated communication: An analysis of task performance and dialogue structure. *Applied Cognitive Psychology*, 17, 325–348. doi:10.1002/acp.868
- Nickerson, R. S. (1985). Understanding understanding. *American Journal of Education*, 93(1), 201–239.
- Nivre, J., Allwood, J., & Ahls n, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9(1), 1–26.
- Nivre, J., Allwood, J., Gr nqvist, L., Gunnarsson, M., Ahls n, E., Vappula, H., Hagman, J., Larsson, S., Sofkova, S., & Ottensj , C. (2004). *G teborg transcription standard version 6.4*. Department of Linguistics, G teborg University.
- N th, E., Batliner, A., Warnke, V., Haas, J., Boros, M., Buckow, J., Huber, R., Gallwitz, F., Nutt, M., & Niemann, H. (2002). On the use of prosody in automatic dialogue understanding. *Speech Communication*, 36, 45–62.
- Olson, G., & Olson, J. (2000). Distance matters. *Human Computer Interaction*, 15, 139–178.
- Orestr m, B. (1983). *Turn-taking in English conversation*. Lund: Gleerup.
- Paggio, P., & Navarretta, C. (2011). Feedback and gestural behaviour in a conversational corpus of Danish. In P. Paggio, E. Ahls n, J. Allwood, K. Jokinen, & C. Navarretta (Eds.), *Proceedings of the 3rd Nordic Symposium on Multimodal Communication* (pp. 33–39), Northern European Association for Language Technology (NEALT).
- Paggio, P., & Navarretta, C. (2013). Head movements, facial expressions and feedback in conversations: Empirical evidence from danish multimodal data. *Journal on Multimodal User Interfaces: Special Issue on Multimodal Corpora*, 7(1–2), 29–37.

- Palmer, R. E. (1969). *Hermeneutics: Interpretation theory in Schleiermacher, Dilthey, Heidegger, and Gadamer*. Northwestern University Studies in Phenomenology & Existential Philosophy. Northwestern University Press.
- Patel, R., & Grigos, M. (2006). Acoustic characterization of the question-statement contrast in 4, 7, and 11-year old children. *Speech Communication, 48*(10), 1308–1318.
- Paul, S., Seetharaman, P., Samarah, I., & Mykytyn, P. P. (2004). Impact of heterogeneity and collaborative conflict management style on the performance of synchronous global virtual teams. *Information and Management, 41*(3), 303–321. doi:10.1016/S0378-7206(03)00076-4
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation, 4*, 203–228.
- Pollack, I., Rubenstein, H., & Horowitz, A. (1960). Communication of verbal modes of expression. *Language and Speech, 3*, 121–130.
- Pollack, M. E. (1990). Plans as complex mental attitudes. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 77–103). Cambridge & MA: MIT Press.
- Popescu, A. D., Aldea, C. C., & Draghici, A. (2012). Misunderstandings in collaborative engineering teams- a way for improving knowledge sharing and interaction. *Managerial Challenges of the Contemporary Society, 4*, 106–109.
- Popper, K. R. (1957). *The poverty of historicism* (7th ed.). London: Routledge & Kegan Paul 1972.
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*(2), 160–174.
- Pride, J. B. (1985). *Cross-cultural encounters: Communication and miscommunication*. Melbourne: River Seine.
- Prieto, P., Borrás-Comes, J., Tubau, S., & Espinal, M. T. (2013). Prosody and gesture constrain the interpretation of double negation. *Lingua, 131*, 136–150.
- Randolph, J. J. (2005). Free-marginal multirater kappa: An alternative to Fleiss' fixed-marginal multirater kappa. In *Proceeding of Joensuu University Learning and Instruction Symposium 2005*, Joensuu, Finland.
- Rapp, B., & Jackson, P. J. (Eds.). (2003). *Organization and work beyond 2000*. New York: Springer-Verlag.
- Romero-Trillo, J. (2010). Pragmatic triangulation and misunderstanding: A prosodic perspective. *Linguistics and the Human Sciences, 6*(1–3), 43–55.
- Rommetveit, R. (1974). *On message structure*. London: Wiley.
- Roulet, E. (1992). On the structure of conversation as negotiation. In H. Parret, & J. Verschueren (Eds.), (On) Searle on conversation (pp. 91–100). Amsterdam: John Benjamins.
- Ruffner, M., & Burgoon, M. (1981). *Interpersonal communication*. New York: Holt, Rinehart & Winston Publishers.
- Ryan, W. E., & Conover, T. E. (2004). *Graphic communications today* (4th ed.). New York: Thomson, Delmar Learning.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language 50*(4), 696–735.
- Säljö, R. (1982). *Learning and understanding: A study of differences in constructing meaning from a text*. Göteborg Studies in Educational Sciences, Vol 41. Acta Universitatis Gothoburgensis.
- Samovar, L., Porter, R. E., & McDaniel, E. R. (2012). *Intercultural communication: A reader*. (13th ed.). Boston: Wadsworth Cengage Learning.
- Sarangi, S. (1994). Intercultural or not? Beyond celebration of cultural differences in miscommunication analysis. *Pragmatics, 4*(3), 409–427.
- Sarangi, S. & Roberts, C. (Eds.) (1999). *Talk, work and institutional order. discourse in medical, mediation and management settings*. Berlin/New York: Mouton de Gruyter.
- Saussure, de F. (1916). *Course in General Linguistics*. New York: Mc Graw Hill.
- Sayer, I. M. (2013). Misunderstanding and language comprehension. *Procedia-Social and Behavioral Sciences, 70*, 738–748.
- Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.). *Analyzing discourse: Text and talk* (pp. 71–93). Washington, D.C., USA: Georgetown University Press.
- Schegloff, E. A. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics, 25*, 201–218.
- Schegloff, E. A. (1991). Conversation analysis and socially shared cognition. In: L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 150–171). Washington, DC: American Psychological Association.

- Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Psychology*, 97(5), 1295–1345.
- Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 52–133). Cambridge: Cambridge University Press.
- Schegloff, E. A. (1998). Reflections on studying prosody in talk-in-interaction. *Language and Speech*, 41, 235–263.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge, England: Cambridge University Press.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361–382.
- Schul, P. L., & Lamb, C. W. J. (1982). Decoding nonverbal and vocal communications: A laboratory study. *Journal of the Academy of Marketing Science*, 10, 154–164.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Searle, J. R. (1975). A taxonomy of illocutionary acts. In: K. Gunderson (Ed.), *Language, mind, and knowledge: Minneapolis studies in the philosophy of science*, vol. 7 (pp. 344–369). University of Minneapolis Press.
- Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.
- Siegel, S., & Castellan, J. (1988). *Nonparametric statistics for the social sciences* (2nd ed.). New York: McGraw-Hill.
- Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Shin, Y. (2005). Conflict resolution in virtual teams. *Organizational Dynamics*, 34(4), 331–345.  
doi:10.1016/j.orgdyn.2005.08.002
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.
- Sins, P. H. M., Savelsbergh, E. R., van Joolingen, W. R., & van Hout-Wolters, B. H. A. M. (2011). Effects of face-to-face versus chat communication on performance in a collaborative inquiry modeling task. *Computers & Education*, 56, 379–387. doi:10.1016/j.compedu.2010.08.022
- Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *The Journal of the Acoustical Society of America*, 117, 305–318.
- Smith, P. (2004). *Cultural theory: An introduction*. Oxford: Blackwell Publishing.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Blackwell, Oxford and Harvard Up.
- Stacks, D. W., & Salwen, M. B. (Eds.). (2009). *An integrated approach to communication theory and research* (2nd ed.). New York: Routledge Communication Series.
- Steinberg (2002). WaveLab (Version 4.00c). [Computer software]. Hamburg: Steinberg.
- Stone, N. J., & Posey, M. (2008). Understanding coordination in computer-mediated versus face-to-face groups. *Computers in Human Behavior*, 24, 827–851. doi:10.1016/j.chb.2007.02.014
- Straus, S. G. (1997). Technology, group process, and group outcomes: testing the connections in computer-mediated and face-to-face groups. *Human-Computer Interaction*, 12, 227–266.  
doi:10.1207/s15327051hci1203\_1
- Straus, S. G., & McGrath, J. E. (1994). Does the medium matter: The interaction of task type and technology on group performance and member reactions. *Journal of Applied Psychology*, 79, 87–97.
- Svennevig, J. (1999). *Getting acquainted in conversation*. Amsterdam: John Benjamins Publishing Company.
- Tannen, D. (1990). *You just don't understand: Women and men in conversation*. New York: William Morrow.
- Tannen, D. (Ed.) (1993). *Framing in discourse*. New York and Oxford: Oxford University Press.
- Taylor, T. (1992). *Mutual misunderstanding: Scepticism and the theorizing of language and interpretation*. Durham & London: Duke University Press.
- Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4(2), 91–112.
- Thompson, L. F., & Coovert, M. D. (2006). Understanding and developing virtual computer-supported cooperative work teams. In C. Bowers, E. Salas, & F. Jentsch (Eds.), *Creating high-tech teams: Practical guidance on work performance and technology* (pp. 213–242). Washington, DC: American Psychological Association.  
doi:10.1037/11263-010
- Ting-Toomey, S. (1999). *Communicating across cultures*. New York: Guildford Press.
- Tomlinson Jr, J. M., & Fox Tree, J. E. (2011). Listeners' comprehension of uptalk in spontaneous speech. *Cognition*, 119(1), 58–69.

- Tronnier, M., & Allwood, J. (2004). Fundamental frequency in feedback words in Swedish. In *Proceedings of the 18th International Congress on Acoustics (ICA 2004)* (pp. 2239–2242), Kyoto, Japan.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Verdonik, D. (2010). Between understanding and misunderstanding. *Journal of Pragmatics*, 42(5), 1364–1379.
- Vygotsky, L. S. (1962). *Thought and language*. (Haufmann, E., & Vakar, G., Trans.). Cambridge: MIT Press.
- Ward, N. (1996). Using prosodic cues to decide when to produce backchannel utterances. *Spoken Language*, Proceedings of ICSLP, Vol. 3 (pp. 1728–1731), Philadelphia, PA.
- Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 32, 1177–1207.
- Warrens, M. J. (2010). Inequalities between multi-rater kappas. *Advances in Data Analysis and Classification*, 4(4), 271–286.
- Weber, E. G. (1993). *Varieties of Questions in English Conversation: Volume 3 of Studies in discourse and grammar*. Amsterdam & Philadelphia: John Benjamins Publishing.
- Weigand, E. (1999). Misunderstanding: The standard case. *Journal of Pragmatics*, 31(6), 763–785.
- Weigand, E. (2010). *Dialogue: The mixed game*. Amsterdam: John Benjamins.
- Weizman, E. (1999). Building true understanding via apparent miscommunication: A case study. *Journal of Pragmatics*, 31, 837–856.
- Whittaker, S. (2003). Theories and methods in mediated communication. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 243–286). Lawrence Erlbaum Associates Inc.
- Wiener, N. (1948). *Cybernetics or control and communication in the animal and the machine*. Cambridge, MA: MIT Press.
- Wierzbicka, A. (1992). *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations*. Oxford: Oxford University Press.
- Wierzbicka, A. (2010). Cross-cultural communication and miscommunication: The role of cultural keywords. *Intercultural Pragmatics*, 7(1), 1–23.
- Willis, W. D., & Coggeshall, R. E. (2004). *Sensory mechanisms of the spinal cord: Primary afferent neurons and the spinal dorsal horn* (3rd ed.). New York: Kluwer Academic/ Plenum Publishers.
- Wittgenstein, L. (1958). *Philosophical investigations* (2nd ed.). (Anscombe, G. E. M., Trans.). Oxford: Basil Blackwell.
- Wood, J. T. (2011). *Communication mosaics: An introduction to the field of communication* (6th ed.). Boston: Wadsworth Cengage Learning.
- Wright, K. B., & Webb, L. M. (Eds.). (2011). *Computer-mediated communication in personal relationships*. New York: Peter Lang.
- Xu, Y., & Wang, M. (2009). Organizing syllables into groups: Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics*, 37, 502–520.
- Yngve, V. H. (1970). On getting a word in edgewise. In M. A. Campell et al. (Eds.), *Papers from the sixth regional meeting of the Chicago linguistic society* (pp. 567–577). Chicago: Chicago Linguistic Society.
- Zaefferer, D. (1977). Understanding misunderstanding: A proposal for an explanation of reading choices. *Journal of Pragmatics*, 1(4), 329–346.
- Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, 38(3), 553–586. doi:10.1017/S027226311500025X
- Ziff, P. (1972). *Understanding understanding*. Ithaca: Cornell University Press.
- Zlatev, J. (2009). Levels of meaning, embodiment, and communication. *Cybernetics and Human Knowing*, 16(3–4), 149–174.
- Zofi, Y. (2011). *A manager's guide to virtual teams*. New York: AMACOM.
- Zuraidah, M. D., & Knowles, G. (2006). Prosody and turn-taking in Malay broadcast interviews. *Journal of Pragmatics*, 38, 490–512.

# Appendix A:

## Transcription conventions

The empirical interaction data was transcribed and checked according to The Göteborg Transcription Standard (GTS) version 6.4 (Nivre et al., 2004).

The main transcription conventions used in this thesis are:

< >	indicates the scope of a comment, number identifies a corresponding comment
@	initiates the corresponding comment
/, //, ///	the number of slashes indicate length of a pause: a short pause (/) has a duration of the same order of magnitude as a word (given the current speech rate); a long pause (//) has a duration of several seconds and is noticeable as a "gap" in the speech flow; when in doubt, mark a pause as intermediate (/).
	a vertical bar indicates silence: silences are not pauses but simply that time passes without anybody saying anything
<   >	indicates a pause where communicative gestures are inserted
{ }	contains letters of the written form that are not pronounced in the spoken form
[1 bla ]1	indicates overlapping speech in a numbered order
:	indicates prolongation of a sound
( bla )	indicates an uncertain utterance
(...)	indicates non-audible speech
\$	identifies a speaker
Sm1/2	Swedish male 1/2
Sf1/2	Swedish female 1/2
Cm1/2	Chinese male 1/2
Cf1/2	Chinese female 1/2

# Appendix B:

## Coding schemes

A variant of the MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena (shortened as MUMIN) (Allwood et al., 2007) was used in this thesis. It includes the classifications of gestural and vocal-verbal micro-feedback (as presented in Section 4.9) and also a coding scheme in primary for the communicative functions of micro-feedback.

Communicative functions of micro-feedback:

FB	micro-feedback
VFB(E)	vocal-verbal micro-feedback (eliciting)
GFB(E)	gestural micro-feedback (eliciting)
CPUE/A	contact, perception, understanding, emotion/attitude
P	(sufficiently) correct perception
misP	misperceived or incorrectly perceived
-P	cannot or do not perceive
U	sufficient understanding
misU	misunderstanding
-U	non-understanding

# Appendix C:

## The reading material for Study 2

Chapter 3 extracted from book *Leadership in a Diverse and Multicultural Environment: Developing Awareness, Knowledge, and Skills* (Connerley & Pedersen, 2005)

### 3

---

## *Cultural Frameworks and Their Importance for Leaders*

---

### Major Objective

To identify and describe models of culture and their relationship to leader actions in the workplace

### Secondary Objectives

1. To identify several models of the various dimensions of culture
  2. To discuss intercultural sensitivity
  3. To outline a three-stage developmental sequence for developing multicultural awareness, knowledge, and skills
- 

Culture has been defined as the source of ties that bind members of societies through an elusive “socially constructed constellation consisting of such things as practices, competencies, ideas, schemas, symbols, values, norms, institutions, goals, constitutive rules, artifacts, and modifications of the

physical environment” (Fiske, 2002, p. 85). These internalized rules create traditions that often go deeper than reason (Stuart, 2004). Using Kelly’s (1955) terms, cultural orientation could be thought of as the master plan behind superordinating constructs that covertly influence manifest cognitive content.

Because much of the strength of cultural influences stems from the fact that they operate in the background of behavior at the value, linguistic, and construct levels, people often have difficulty defining their cultural influences, and social scientists have difficulty measuring them. (Stuart, 2004, p. 4)

This chapter identifies several models of culture before introducing the three-stage developmental sequence that will be used as the foundation for the rest of the book.

## **Leaders and Culture**

The culture that we are embedded in inevitably influences our views about leadership (Hofstede, 1993). To make sense of the different types of cultural influence, Gardenswartz, Rowe, Digh, and Bennett (2003) developed the three cultures model, which posits three cultural influences at work in corporations: personal culture, national culture, and organizational culture. They state that the model is based on work in global corporations, but it is our premise that it captures cultural influences in both global and non-global corporations.

Personal culture is the shared combination of an individual’s traits, skills, and personality formed within the context of his or her ethnic, racial, familial, and educational environments. Every one has a unique personal culture.

National culture is a shared understanding that comes from the combination of beliefs, values, attitudes, and behaviors that have provided the foundation for the heritage of a country. Although national culture is a shared understanding, as is well known, individuals within a nation still have a very wide range of beliefs about their nation.

Corporate culture is a combination of widely shared institutional beliefs, values, and the organization’s guiding philosophy that is usually stated in its vision, mission, and values statements (Gardenswartz et al., 2003). Similar to national culture, individuals within an organization often view their organization differently. These varying views often align themselves with individuals’ levels within the company hierarchy. This results in leaders often having different views about their corporate culture compared to those in the lower levels in the organization. Keeping in touch with how these views differ is an important part of every leader’s job.

For interactions within organizations, culture is a mix of personal, national, and corporate culture. The focus of this book is on the personal

culture that has developed within the national culture that takes place within the corporate culture. Culture is not external but is “within the person”; it is not separate from other learned competencies. Developing multicultural awareness, knowledge, and skills should be seen as a professional obligation as well as an opportunity for a leader. With the millions of employees living and working in diverse environments, there are abundant opportunities for enhancing multicultural awareness, knowledge, and skills. People who live in an unfamiliar culture are likely to become more multicultural in their awareness of alternative values, habits, customs, and lifestyles that were initially strange and unfamiliar. Sometimes they have learned to adjust even more profoundly and effectively than they themselves realize. They have learned to respond in unique ways to previously unfamiliar situations and come up with the right answers without always being aware of their own adjustment process. Again, as stated earlier, given demographic changes, understanding culture has great implications both domestically and internationally.

### **Seminal Work on Culture**

Differences in culture can significantly affect leadership practices. One of the earliest identifications of the dimensions of culture was developed by Kluckhohn and Strodtbeck (1961).

- Basic nature of human beings: Good—left to their own devices, individuals are basically good and will act in a reasonable and responsible manner; Evil—individuals are basically evil and are not to be trusted; Mixed—individuals are a mixture of good and evil.
- Relationships among people: Individualistic—the primary responsibility of an individual is to him- or herself. Individual abilities and characteristics are the primary consideration; Group—responsibility to family and groups is most important. Ability to fit into the group is more important than individual ability; Hierarchical—Similar to the group orientation with the addition that distinct differences in status are expected and respected.
- Activity orientation: Being—the point of life is to live and experience an understanding. Activity for activity’s sake is unimportant. Doing—the point of life is actually to do things, be involved, and accomplish goals.
- Relation to nature: Subjugation—nature and the environment determine human activities; Harmony—humans should live in harmony with their environment; Domination—Humans can exert domination over their environment while they control their own destinies.

- Time orientation: Past—history is important in determining our present actions; Present—the current situation should determine what we do as we focus our energy on the present; Future—our actions should concentrate on the future and the attainment of future goals.

The power of national cultures can also be understood by examining seminal research conducted by Geert Hofstede (1984, 1985, 2001). Hofstede conducted research on IBM employees in 40 countries and discovered that cultural values strongly influenced relationships both within and between organizational divisions. Four of the significant cultural dimensions that Hofstede defined have been examined by many researchers. Understanding the way these dimensions influence culture is of increasing importance for both global leaders and those managing a diverse workforce.

- Power Distance refers to whether individuals accept inequality in power, including within an organization. Low power distance means individuals expect equality in power and do not accept a leader's authority just because of the leader's position.

- Uncertainty Avoidance refers to the feeling of comfort or discomfort associated with levels of uncertainty and ambiguity. Low uncertainty avoidance means that individuals easily tolerate unstructured and unpredictable situations.

- Individualism and Collectivism refer to the social frameworks in which individuals prioritize individual or group needs. In Individualistic societies, individuals are expected to take care of themselves; in collectivistic societies, individuals are expected to look out for one another, and organizations protect their employees' interests.

- Masculinity and Femininity refer to the emphasis a culture places on emotional and social roles and work goals. A masculine culture reflects a preference for assertiveness, achievement, and material success. A feminine culture values relationships, cooperation, and quality of life. Despite the label for this dimension, both men and women subscribe to the dominant value, whether it is masculine or feminine.

### **Additional Important Research Related to Culture**

It must be noted that although Hofstede's landmark work is widely cited, it also has its critics (see Dickson, Hanges, & Lord, 2001, for a review of literature defining culture, including a section on criticisms of Hofstede's work). In

addition to the work done by Hofstede, many other influences on behavior have been classified. Trompenaars and Hampden-Turner (1998) also focus on cultural differences and how they affect business and management. They present data from more than 30,000 participants of training programs and describe seven dimensions of cultural difference:

- **Universalism versus Particularism:** In a Universalist culture, rules are more important than relationships; legal contracts are drawn up and are seen as trustworthy, you must honor them; In a Particularist culture, whether a rule applies “depends” on the situation and relationships evolve.
- **Individualism versus Communitarianism:** Essentially the same as Hofstede’s Individualism versus Collectivism dimension.
- **Neutral versus Affective (Emotional):** Individuals in a Neutral culture hide their thoughts and feelings while maintaining a cool self-control. Speech is often monotone, and individuals do not touch each other. In an Affective culture, individuals express their thoughts openly while using gestures and dramatic expressions. There is often a great deal of passion in discussions, and individuals often touch.
- **Specific versus Diffuse:** In Specific cultures, individuals are direct, clear, blunt, and to the point while examining the facts. In Diffuse cultures, individuals are more indirect and tactful. The context of a situation matters, and they tolerate ambiguity.
- **Achievement versus Ascription:** In Achievement-oriented societies, there is little focus on titles, which are used only when they reflect competencies. Leaders are judged on what they do and know. In Ascribed-status societies, titles are important; the boss is the boss, regardless of the situation. Leaders with authority are usually older males.
- **Attitudes toward Time: Past versus Present versus Future—**Essentially the same as Kluckhohn and Strodtbeck’s time orientation dimension.
- **Internal versus External control:** Essentially the same as Kluckhohn and Strodtbeck’s subjugation and domination orientations in the sense of being able or not being able to control what happens in the environment.

Ten distinct types of motivational values have been derived from the universal requirements of human existence and verified in cross-cultural research by Schwartz (1992) and colleagues (e.g., Sagiv & Schwartz, 1995; Schwartz & Bilisky, 1987, 1990; Schwartz & Huisman, 1995; Schwartz & Sagiv, 1995).

The 10 types of motivational values are as follows:

Power: Social status and prestige, dominance or control over people and resources

Achievement: Demonstrating competence according to social standards

Stimulation: Challenge, excitement, and novelty in life

Self-Direction: Independent thought and action

Hedonism: Pleasure and sensuous gratification for oneself

Security: Harmony, stability, and safety of society, relationships, and self

Conformity: Restraining actions or impulses that would likely upset or harm others and violate social expectations

Tradition: Commitment, respect, and acceptance of the ideas and customs that traditional culture and religion provide

Benevolence: Preserving and enhancing the welfare of all people with whom one is frequently in contact

Universalism: Being broadminded and having an appreciation, understanding, and tolerance for the welfare of all people and for nature

Understanding values is important in cross-cultural interactions. Research has shown that personal values accounted for a large proportion of individual variation in readiness for contact with others from a different group (Sagiv & Schwartz, 1995).

Ronen and Shenkar (1985) clustered countries based on patterns of similarity in employees' attitudes toward work and how well it met their needs. Eight country clusters, with four countries remaining independent of any cluster, were identified. The clusters include Near Eastern, Arab, Far Eastern, Latin American, Latin European, Anglo, Germanic, and Nordic. As an example, the Anglo cluster is made up of the United States, Canada, Australia, New Zealand, the United Kingdom, Ireland, and South Africa. The four independent countries are Brazil, Japan, India, and Israel, which have unique religions, languages, and/or histories. Cluster classifications were made after a comprehensive review of previous research that included assessments of how thousands of employees in close to 50 countries responded to questions about the importance of various work goals, the extent to which work satisfies certain needs, organizational and managerial issues, and the nature of work roles and interpersonal relationships (e.g., how well managers relate to subordinates).

Although there are limitations to Ronen and Shenkar's approach (many countries are not included), leaders can use the clusters to determine where broad similarities and differences of values and attitudes may exist between the countries that are listed. Since business practices often reflect values and attitudes, this can help leaders to be more effective in their interaction with those from cultures not similar to their own.

# Appendix D: General personal information questionnaire

Name	<input type="text"/>
Age	<input type="text"/>
Gender	<input type="text"/>
Nationality	<input type="text"/>
Which country did you grow up in	<input type="text"/>
First language (mother tongue)	<input type="text"/>
Current school or university	<input type="text"/>
Which study program of which level	<input type="text"/>
Which languages do you usually use in daily communication?	<input type="text"/>
Which language do you use most frequently?	<input type="text"/>
Email address	<input type="text"/>

# Appendix E: Follow-up interview

## Interview Questions

1. How did you understand the two tasks? What were the two tasks?
2. Which did you pick as the three differences and three similarities between Sweden and China? Which did you pick as the most obvious difference and similarity? Why did you pick them? Do you think your partner had the same opinion?
3. How did you feel about your face-to-face communication with your partner? How did you feel about your video-mediated communication?
4. How different did you experience the two (face-to-face and video-mediated) communications? For instance, in which way are they different? How do you think of these differences? What did you enjoy and what didn't you? Do you have any suggestion on how to improve your experience in the activity?
5. How experienced are you with video-mediated communication? In what kind of context or situation or activity do you think it works well for you, and in what kind of context or situation or activity it doesn't? Can you give examples?
6. Before you participated in this project, did you have any professional or specialised education or work experience with culture and communication? If yes, how much or to which level did you know?
7. After all, what have you learned anything, from the background material, the conversations, and discussions? Can you give examples?
8. Do you have any question about the study or anything you are concerned? Please contact me (you have my email [jia.lu@gu.se](mailto:jia.lu@gu.se)), if you have any questions in the future.

# Appendix F:

## Questionnaire regarding the communication experience

### Debriefing Questions

Please rate your experience of the **FACE-TO-FACE situation only**

Ease of communication

very difficult very easy

Pleasantness of communication

very unpleasant very pleasant

Experience of miscommunication

never very often

Ease of understanding the other person

very difficult very easy

Ease of communicating what you wanted

very difficult very easy

Mutual understanding

very low very high

Efficiency (least amount of time and resources consumed with most productivity achieved)

very low very high

Comments about communicating in the face-to-face situation:

Please rate your experience of the **VIDEO-MEDIATED** situation only

Ease of communication	<input type="radio"/>
	very difficult <span style="float: right;">very easy</span>
Pleasantness of communication	<input type="radio"/>
	very unpleasant <span style="float: right;">very pleasant</span>
Experience of miscommunication	<input type="radio"/>
	never <span style="float: right;">very often</span>
Ease of understanding the other person	<input type="radio"/>
	very difficult <span style="float: right;">very easy</span>
Ease of communicating what you wanted	<input type="radio"/>
	very difficult <span style="float: right;">very easy</span>
Mutual understanding	<input type="radio"/>
	very low <span style="float: right;">very high</span>
Efficiency (least amount of time and resources consumed with most productivity achieved)	<input type="radio"/>
	very low <span style="float: right;">very high</span>

Comments about communicating in the video-mediated situation:

**Which mode would you choose in the future for the following situations?**

	Face-to-face	Video-mediated
Collaborative learning activities	<input type="radio"/>	<input type="radio"/>
Communication with a stranger for personal purposes (such as dating)	<input type="radio"/>	<input type="radio"/>
Communication with a friend for social contact	<input type="radio"/>	<input type="radio"/>
Communication with your family	<input type="radio"/>	<input type="radio"/>
Business communication with someone for the first time	<input type="radio"/>	<input type="radio"/>
Business communication with someone you had earlier contact with	<input type="radio"/>	<input type="radio"/>

# Appendix G: Open discussion questions as the collaborative problem-solving tasks

## Task instruction

### *Discussion question*

With a reference to the provided background material mainly and possibly your own experiences and knowledge also, please pick up three points you and your partner agree upon as *different* between Sweden and China. Then, please decide which is most obvious, and give arguments and examples on why you think so. It is ok, if you cannot agree with each other, but please explain to the other why you cannot agree, how you think, and what your reasons are.

Please complete this task within 8 minutes. The researcher will come and check then.

## Task instruction

### *Discussion question*

With a reference to the provided background material mainly and possibly your own experiences and knowledge also, please pick up three points you and your partner agree upon as *similar* between Sweden and China. Then, please decide which is most obvious, and give arguments and examples on why you think so. It is ok, if you cannot agree with each other, but please explain to the other why you cannot agree, how you think, and what your reasons are.

Please complete this task within 8 minutes. The researcher will come and check then.

# Appendix H: Consent form for participation in the project

CHALMERS



GÖTEBORGS UNIVERSITET

## Consent to participate in the research project *On communication, collaboration, and technology*

I've been informed about the research project on communication, collaboration, and technology in both written and oral means. The project is implemented at the Department of Applied Information Technology at the University of Gothenburg | Chalmers. I am aware of that my participation in the study is voluntary and that I can at any time and without explanation withdraw my participation and even ask the researchers to erase the data on me.

I have been informed that I will be audio and video recorded in the project and that the research group will study the recorded material and use it in an anonymized form in presentations in scientific and educational contexts. I agree to participate in the study.

Signature of the participant: \_\_\_\_\_

Printed name of the participant: \_\_\_\_\_

Place and date: \_\_\_\_\_

### *Assurance from the responsible researcher*

I certify that I have provided information about the study of the above participant. I have gone through and explained the purpose of the research project, how it will be implemented, and how the collected data will be handled. The participant has had the opportunity to ask questions and has received a copy of the project information and the consent form.

Signature of the responsible researcher: \_\_\_\_\_

Printed name of the responsible researcher: \_\_\_\_\_

Place and date: \_\_\_\_\_

INSTITUTIONEN FÖR TILLÄMPAD INFORMATIONSTEKNOLOGI  
Göteborgs universitet | Chalmers  
41296 Göteborg  
Besök: Forskningsgöngen 6, Livsholmen  
Webb: [www.chalmers.se/mci](http://www.chalmers.se/mci)