CONTRIBUTIONS TO
HEARING RESEARCH

Volume 47

Anna Josefine Munch Sørensen

The effects of noise and hearing loss on conversational dynamics

DTU Health Tech
Department of Health Technology

The effects of noise and hearing loss on conversational dynamics

PhD thesis by

Anna Josefine Munch Sørensen

HEARING SYSTEMS

Technical University of Denmark

2021

The defense was held on June 30, 2021.

This PhD dissertation is the result of a research project carried out at the Hearing Systems Section, Department of Health Technology (formerly Department of Electrical Engineering), Technical University of Denmark.

The project was financed by William Demant Foundation.

Supervisors

Assoc. Prof. Bastian Epp
Hearing Systems Group
Department of Health Technology
Technical University of Denmark
Kgs. Lyngby, Denmark

Assoc. Prof. Ewen N. MacDonald
Hearing Systems Group
Department of Health Technology
Technical University of Denmark
Kgs. Lyngby, Denmark

Department of Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada

Senior Scientist Lars Bramsløw
Eriksholm Research Centre
Snekkersten, Denmark

Abstract

Understanding how normal-hearing and hearing-impaired people communicate in their everyday life is an important area of research within hearing diagnostics and hearing aid (HA) development. Traditionally, aspects of speech comprehension and speech production are studied in isolation and results from these studies are generalized to real-world performance. However, the act of conversing is not just the sum of speech comprehension and production. It involves a complex overlap between the two, manifested in a dynamic feedback process between interlocutors, i.e., conversational partners. Thus, there is a need to study people in interaction to understand the impacts of hearing loss and HA amplification on communication. In this thesis, the effects of hearing loss, noise interference, and HA amplification on measures of temporal dynamics and speech production between interlocutors were investigated in remote and face-to-face conversations. It was shown that in the presence of noise compared to quiet, both normal-hearing (NH) and hearing-impaired (HI) interlocutors responded later and with more variability when taking a turn. They also produced longer units of connected speech, i.e., interpausal units (IPUs), and they spoke louder. When conversing with an HI interlocutor in the presence of noise, NH participants altered their conversational dynamics more than when conversing with an NH partner. The NH interlocutors changed their strategy from conversing at negative to positive signal-to-noise ratios (SNRs), and they slowed down their speech rates, which were unaltered in NH/NH conversations. HI interlocutors answered with more variability and produced longer IPUs than their NH partners. When conversing remotely, the HI interlocutors tended to dominate the conversation. When receiving simple HA amplification in face-to-face conversations, the HI participants spoke faster, timed their responses with greater precision, and produced shorter IPUs. These results suggest that the considered outcome measures were sensitive to communication difficulty and the improvements provided by the HA amplification, even in quiet. Therefore, the proposed measures of communication performance have the potential to be used in the development of new experimental paradigms that evaluate how hearing loss affects real-life communication and to assess the benefit of hearing devices in compensating for these deficits.

Resumé

At forstå hvordan normalthørende (NH) og hørehæmmede (HH) kommunikerer i deres hverdag er et vigtigt forskningsområde inden for hørediagnostik og høreapparatsudvikling. Traditionelt undersøges aspekter af taleforståelse og taleproduktion isoleret, og resultaterne fra disse studier generaliseres til, hvordan folk interagerer i den virkelige verden. Imidlertid er det at samtale ikke blot en sum af taleforståelse og taleproduktion; det indebærer et komplekst overlap mellem de to processer i en dynamisk feedbackproces mellem samtalepartnere. Der er således et behov for at studere folk i interaktion for at forstå indvirkningen af høretab og høreapparatsforstærkning på kommunikation.

I denne afhandling blev påvirkningerne af høretab, støjinterferens og høreapparatsforstærkning på målinger af temporal dynamik og taleproduktion mellem samtalepartnere undersøgt, både når samtalepartnerne var separeret og konverserede ansigt til ansigt. Det blev vist, at når NH og HH samtalepartnere kommunikerede i baggrundsstøj sammenlignet med i stilhed, svarede de begge senere og med mere variation, når de tog deres tur. De producerede også længere taleenheder (sammenhængende enheder af tale omringet af stilhed), og de talte højere. Når de NH konverserede med en HH samtalepartner i støj ændrede de deres samtaledynamik mere end når de konverserede med en NH samtalepartner. De ændrede deres strategi fra at kommunikere i negative til at kommunikere i positive signal/støj-forhold, og de talte langsommere i støj, hvorimod støj ikke påvirkede deres talehastighed i samtaler med NH deltagere. De HH deltagere svarede med mere variation og producerede længere taleenheder end deres NH samtalepartnere. I samtaler hvor de ikke kunne se hinanden havde de HH deltagere en tendens til at dominere samtalen. Når de HH deltagere modtog simpel høreapparatsforstærkning i samtalerne optaget ansigt til ansigt talte de hurtigere, havde en mere præcis timing af deres respons og producerede kortere taleenheder.

Disse resultater indikerer, at disse mål for samtaledynamik var følsomme nok til at kunne detektere kommunikationsproblemer og de forbedringer høreapparatsforstærkningen gav de HH, selv når der ingen baggrundsstøj var. Derfor har disse mål potentialet til at blive brugt til udviklingen af nye eksperimentelle paradigmer, der evaluerer, hvordan høretab påvirker kommunikation i den virkelige verden og til at vurdere i hvilken grad diverse høreapparatsstrategier hjælper med at kompensere for disse påvirkninger.

Acknowledgments

First of all, I would like to thank Adam Weisser for introducing me to the concept of conversational turn-taking when I was an intern at GN ReSound in 2014. I am thankful for the many thought-expanding conversations about the relevance of turn-taking in hearing research and how it could be utilized to develop our understanding of the difficulties people experience during everyday conversation in the presence of noise and/or when people have a hearing impairment. I would also like to thank Ewen MacDonald for embarking on this journey with me, being my supervisor both for my B.Eng., my M.Sc., and my Ph.D. Thank you for all your help in the lab and the many constructive discussions. Thanks to Bastian Epp for the great discussions and for always finding a solution to whatever problem may arise. I would like to thank the rest of my former and current supervisor team, Thomas Lunner, and Lars Bramsløw, for your support and valuable discussions. Thanks to Eline Borch Petersen for great collaborations in both supervising students and running our own project. Thanks to my students who had the courage to investigate conversational dynamics with me, providing me invaluable insights through your studies. Thank you to all my great colleagues at Hearing Systems. I have truly enjoyed being around so many incredibly smart, helpful and brilliant people. I would also like to extend my thanks to Facebook Reality Labs and in particular Owen Brimijoin for giving me the opportunity to work with you for half a year, having fruitful and encouraging discussions about conversational turn-taking, and expanding my horizon on the perspectives of the work I am doing. Thanks to Lindsey Kishline for your love and support throughout this process.

Thanks to all the people who participated in my experiments. Without you there would not have been any thesis. And, of course, a big thanks to the William Demant Foundation for funding this project.

A special thanks to my family. To my beloved husband Casper for making this an exciting journey, for supporting me throughout its ups and downs, for the many encouraging and supportive conversations, and for moving with me to Redmond when the opportunity arose. To my sister Thilde for supporting and understanding me, and putting everything into perspective. To her and my nephews, Eddie and Loui, for keeping my spirit up high and making me laugh like no one else can. To my sister-in-law, Line, for drawing such a nice cover for my thesis, you are so talented. To my brother-in-law, Jonas, my parents-in-law, Annette and Bent, and my aunt, Lene, for your support and for believing in me. And finally, thanks to my parents, whom are always with me in spirit.

Related publications

Journal papers

- Sørensen, A. J. M, Fereczkowski, M., & MacDonald, E. (**In press**). "The effects of noise and second language on conversational dynamics in task dialogue," Trends in Hearing. DOI: 10.1177/23312165211024482.
- Sørensen, A. J. M., MacDonald, E., & Lunner, T. (**Submitted**). "Conversational dynamics in task dialogue between normal-hearing and hearing-impaired interlocutors," submitted for publication.
- Borch, E., MacDonald, E., & Sørensen, A. J. M. (**Submitted**). "The effects of hearing aid amplification and noise on conversational dynamics between normal-hearing and hearing-impaired talkers," submitted for publication.

Conference papers

- Sørensen, A. J. M., Fereczkowski, M., & MacDonald, E. (**2020**). "Effects of noise and L2 on the timing of turn taking in conversation," Proceedings of the International Symposium on Auditory and Audiological Research, 7, 85-92.
- Sørensen, A. J. M., MacDonald, E., & Lunner, T. (**2020**). "Timing of turn taking between normal-hearing and hearing-impaired interlocutors," Proceedings of the International Symposium on Auditory and Audiological Research, 7, 37-44.
- Watson, S., Sørensen, A. J. M., & MacDonald, E. (**2020**). "The effect of conversational task on turn taking in dialogue," Proceedings of the International Symposium on Auditory and Audiological Research, 7, 61-68.

Published abstracts

- Sørensen, A. J. M., Fereczkowski, M., & MacDonald, E. (**2017**). The effect of noise and second language on turn-taking in task-oriented dialogue. Cognitive Hearing Science for Communication (CHSCOM) 2017.

- Sørensen, A. J. M., Fereczkowski, M., & MacDonald, E. (**2017**). The effect of noise and second language on turn taking in task-oriented dialog. The Journal of the Acoustical Society of America 141, 3520 (2017). 10.1121/1.4987409.

- Sørensen, A. J. M., MacDonald, E., & Lunner, T. (**2019**). Conversational dynamics between normal-hearing and hearing-impaired interlocutors. Cognitive Hearing Science for Communication (CHSCOM) 2019.

- Sørensen, A. J. M., MacDonald, E., & Lunner, T. (**2019**). Task dialogue between normal-hearing and hearing-impaired talkers in quiet and noise. 11th Speech in Noise Workshop, Ghent, Belgium, January 2019.

- Sørensen, A. J. M., & Brimijoin, O. W. (**2020**). The development of a multimodal communication behaviour capture system. 12th Speech in Noise Workshop, Toulouse, France, January 2020.

Data and task material

- Sørensen, A. J. M. (**2021**). "PDF and PSD files of DiapixDK picture materials - adapted Danish version [Data set]," Zenodo. 10.5281/zenodo.4638693.

- Sørensen, A. J. M, Fereczkowski, M., & MacDonald, E. N. (**2020**). "Task dialog by native-Danish talkers in Danish and English in both quiet and noise," Zenodo. 10.5281/zenodo.1204951.

Contents

List of figures

List of acronyms

EEG: Electroencephalogram

ERP: Event-Related Potential

FTO: Floor-Transfer Offset

HA: Hearing Aid

HAD: Hearing Assistive Device

HI: Hearing Impaired

IPU: Interpausal Unit

L1: First Language

L2: Second Language

NH: Normal Hearing

NHR: Noise-to-Harmonic Ratio

RMS: Root-Mean-Square

SLM: Sound Level Meter

SNR: Signal-to-Noise Ratio

SPL: Sound Pressure Level

SRT: Speech Reception Threshold

VAD: Voice Activity Detection

1 General introduction

Communication is an integral part of human social interaction. One of the most detrimental consequences of having a hearing impairment is that people tend to withdraw from and avoid social situations (for a review, see Palmer et al., 2016), especially if there is background noise, as their ability to communicate is significantly reduced, and the effort they have to spend to maintain a conversation is increased (Beechey et al., 2020b; Kiessling et al., 2003). Because of these increased communication challenges for hearing-impaired (HI) individuals, many tend to withdraw from social interactions, which can negatively impact people's social lives and health (Palmer et al., 2016). Therefore, one of the most important goals of hearing rehabilitation is to regain HI individuals' ability to participate in spoken interaction. The traditional way of assessing the performance of hearing aids is by evaluating the speech understanding or enhancement for different processing strategies. However, conversation involves an overlap between speech comprehension and production in a dynamic feedback process between two or more people (e.g., Donnarumma et al., 2017; Garrod and Pickering, 2004; Levinson and Torreira, 2015). The purpose of this thesis was to investigate whether conversational dynamics could be used as objective measures to assess people's difficulty participating in conversation.

1.1 Conversation in hearing science

The strive for ecological validity is a growing topic in hearing science. In their consensus paper, Keidser et al. (2020) defined the concept: "*In hearing science, ecological validity refers to the degree to which research findings reflect real-life hearing-related function, activity, or participation.*" In their paper, they underlined the general lack of understanding of the dynamic processes between interlocutors (conversational partners) when engaging in conversation and how they are affected when people have a hearing loss.

Traditionally, hearing ability and hearing rehabilitation outcomes using hearing assistive devices have been studied through measures of speech comprehension or hearing sensitivity. The tradition in hearing science is to have a reductionist approach to experimental design in which all other variables but the independent variable should be controlled for. The strength of this approach is that any changes in the outcome variable(s) can be accounted for by the manipulations to the independent variable(s). However, the drawback of this approach is that the effects observed may be isolated to that experimental setup, and thus the results may not generalize to realistic scenarios. On the other hand, conducting experiments in realistic scenarios poses the risk of including too much variability, making the results less reproducible, and making it more difficult to attribute changes in the outcome variable(s) to the independent variable(s). Thus, there is a trade-off between controllability and realism in hearing science experiments (Keidser et al., 2020; Mansour, 2021).

The consensus paper by Kiessling et al. (2003) outlines the differences and overlaps between *hearing*, *listening*, *comprehending* and *communicating*. *Hearing* is a passive function involving the detection and discrimination of sounds. *Listening* involves actively attending to one source and ignoring others, which is mentally effortful. *Comprehension* is the process of inferring meaning or intent from a signal, e.g., a speech signal. If a signal is degraded, there is a higher risk of misperceiving it, in which case a person can use higher-level processing to infer meaning from the context. *Communication* involves an overlap between hearing, listening, comprehension and production. Much research has gone into understanding the role of hearing and listening, while recently, comprehension studies have grown more popular with advances within cognitive hearing science. Often, studies of hearing, listening, and comprehension can be highly controlled, as participants' behavior does not depend on a feedback loop with an interlocutor, and thus it allows one to change one parameter at a time in an experimental setup and only influence the listener. However, communication is a dynamic, interactive feedback process between interlocutors involving context and the use and development of cognitive models of one's interlocutor (Carlile and Keidser, 2020). Making adjustments to one parameter affects both participants in a conversation. Thus, studies of conversational interaction are inherently less controllable and have not been given much attention in hearing science so far. Typical studies of communication in hearing science have been through interviews or qualitative observations of people in interaction. However, recently some studies have investigated the impacts of hearing loss on conversation from objective observations of people's interaction (Beechey et al., 2020a,b; Hadley et al., 2019, 2020b; Hazan et al., 2018a,b).

1.2 Conversational turn-taking

The fundamental part of having a conversation with another person is the switching of turns between interlocutors. This skill is developed early in life and is a central part in developing social skills. Even from an early age, children aim to respond quickly, and as the complexity of their responses increases, their timing of turns is delayed, and they practice this skill until they reach adult turn-timing (Casillas et al., 2015). The typical response time for adults is around 200 ms with language variations of around 250 ms around this average (Heldner and Edlund, 2010; Levinson and Torreira, 2015; Stivers et al., 2009). Timing a turn within the expected time has shown to be socially important, as a divergence from the expected response time carries meaning by itself (e.g., Mertens and Ruiter, 2021). Extended gaps of 600 ms or more after the offset of a person's turn are associated with dispreferred responses such as misunderstandings or negative responses to invitations (Bögels et al., 2015a; Kendrick and Torreira, 2015), and it has been suggested that extended gaps can signal difficulty in conversation (Mertens and Ruiter, 2021).

When engaging in conversation, people have to simultaneously comprehend and predict the upcoming end of their interlocutor's turn and plan their own response (e.g., Donnarumma et al., 2017; Garrod and Pickering, 2004; Levinson and Torreira, 2015). It is cognitively demanding to perform these processes in parallel as they take up capacity from the same cognitive systems (Donnarumma et al., 2017; Hagoort and Indefrey, 2014; Menenti et al., 2011; Segaert et al., 2012). However, it has been argued that people optimize for fast response times by planning in overlap with their interlocutor's incoming turn at the expense of increased cognitive demands as manifested by increased peak and mean pupil dilation responses, as well as longer peak latencies (Barthel and Sauppe, 2019). People use prediction of the content and end of their interlocutor's turn to recognize their interlocutor's action faster, which in return can free up resources that can be used for planning their utterance. To facilitate rapid turn-taking on the part of the responder, the person producing the current turn can use signal-enhancing strategies to increase the predictability of their utterance (Donnarumma et al., 2017).

1.2.1 How noise and hearing loss could impact content prediction and turn-end prediction cues

In conversation, people predict what their interlocutor is going to say next. They do so by aligning their mental models of the situation by mimicking each other's behavior in a process of *joint action* (Donnarumma et al., 2017; Garrod and Pickering, 2004). The alignment process is implicit and automatic, but when a mismatch is detected between the expected and what is perceived, people recruit explicit cognitive resources to infer meaning from the missing information, which is cognitively demanding. According to the Ease of Language Understanding (ELU) model (Rönnberg et al., 2008, 2013), there is a higher probability for mismatches to occur when the speech is degraded due to hearing loss or noise interference. The presence of a noise masker can lead to a decrease in the resources available for communication, by allocating resources to ignore the masker (Mattys et al., 2009). In the presence of an energetic noise masker, sublexical cues such as acoustic cues may be more affected than lexical-semantic cues. In this case, people would rely more on lexical-semantic cues to predict the content of their interlocutor's speech. However, if the signal is severely acoustically degraded, people may be unable to make use of lexically-driven top-down knowledge, in which case the few glimpses of the acoustic cues in the masker becomes the source of reliance (Mattys et al., 2009).

The cues used for predicting upcoming turn-ends have been widely discussed, and some studies propose that only lexical and syntactic cues are necessary for rapidly producing and timing turns (Corps et al., 2018a; De Ruiter et al., 2006), while other studies suggest the importance of both prosodic and lexicosyntactic cues (Bögels and Torreira, 2015; Brusco et al., 2020; Duncan et al., 1972; Gravano and Hirschberg, 2011; Hjalmarsson, 2011). The following features have been shown to be cues to turn-yielding: a drop in loudness, a rising or falling pitch contour, an increase in vocal jitter, shimmer, and noise-to-harmonic ratio (NHR), a drawl on the final syllable or the stressed syllable of a terminal clause, along with the termination of hand gestures, the use of stereotyped expressions, and the completion of a grammatical clause (Brusco et al., 2020; Duncan et al., 1972; Gravano and Hirschberg, 2011). On the other hand, increased phrase-final lengthening, the use of filled pauses or word fragments, a sustained pitch contour (Gravano, 2010) and the use of hand gestures (Duncan et al., 1972) are turn-holding cues, signaling that the person is not finished talking. The more turn-yielding cues that are present in a turn, the more likely a switch of the conversational floor is, but if any turn-holding cues are present, the probability of a switch of the floor drastically decreases (Duncan et al., 1972; Gravano and Hirschberg, 2011; Hjalmarsson, 2011).

HI individuals experience decreased sensitivity to subtle changes in loudness, and if they have temporal processing deficits they experience decreased sensitivity to subtle pitch changes (Moore, 1987). These cues are important for prosody (Kalathottukaren et al., 2015). HI with profound hearing loss have been shown to be less sensitive to changes in fundamental frequency, F0, and thus less sensitive to changes in intonation and stress patterns (Grant, 1987b), and to have a reduced ability to distinguish between rising and falling intonation (Grant, 1987a). Thus, for HI individuals, perceiving the cues for turn-end prediction may be more difficult, as turn-holding and turn-yielding cues may not be easily distinguishable.

Normal-hearing (NH) listeners have been shown to have reduced intensity discrimination in noise (Schneider and Parker, 1990). Li and Jeng (2011) found that below 0 dB SNR, NH listeners' frequency error, slope error and tracking accuracy of voice pitch deteriorated significantly. Zyl and Hanekom (2011), however, found that prosody recognition was intact for NH individuals even at very low SNRs (-8 dB). They also found that the recognition of sentences and words in sentences decreased with decreasing SNR. Thus, NH individuals could experience decreased sensitivity to turn-end prediction cues involving syntax and loudness changes, but it is less clear whether they would experience decreased sensitivity to changes in pitch-dependent cues.

1.2.2 Signal-enhancement strategies

There are several ways in which a person can make themselves better understood in adverse conditions. People make speech modifications in response to changes in the environment, but also in response to the difficulty their interlocutor is experiencing (for an overview, see Cooke et al., 2014). For example, people have been shown to increase their vocal intensity in response to background noise (e.g., Beechey et al., 2018, 2020b; Pearsons et al., 1977; Weisser and Buchholz, 2019), and to decrease their speech rates when their interlocutor is perceiving a degraded version of their voice signal (Hazan and Baker, 2011) or when conversing in the presence of competing talkers (Aubanel and Cooke, 2013). People increase their fundamental and first formant frequencies (F0 and F1) in response to increasing noise level (e.g., Beechey et al., 2018; Hazan and Baker, 2011), and increase their F1 when conversing with an HI individual (Beechey et al., 2020b), indicating increased vocal effort. Speech modifications have been shown to be enhanced in speech produced with communicative intent rather than speech produced without an addressee (Garnier et al., 2010). Interlocutors have been shown to synchronize on various dimensions of conversational dynamics: they synchronize their turn-taking timing, speech rates, pitch, and speech levels (Giles et al., 1991; Levitan and Hirschberg, 2011; Ten Bosch et al., 2005; Wilson and Wilson, 2005). Hadley et al. (2020a) found that people were better at predicting turn-ends of sentences produced in a manner similar to themselves. Any enhancement of turn-end and content prediction cues and features that can enhance alignment between interlocutors could facilitate smooth turn-taking.

1.2.3 Thesis hypothesis

As outlined above, people may be less able to predict the content of their interlocutor's speech and the end of their turns in the presence of noise or when they have a hearing impairment, which may increase their cognitive load due to the recruitment of explicit resources to infer meaning from the degraded speech input. This, in return, would reduce the person's resources that are available for their speech planning. We hypothesized that we would be able to measure implications of this by observing longer and more variable floor-transfer offsets (FTOs), i.e., the offset from when one person stops talking to the onset of the next person's turn (Aubanel et al., 2011; Donnarumma et al., 2017). Additionally, we expected to observe that interlocutors would use signal-enhancement strategies in the presence of noise and when conversing with an HI individual.

1.3 Overview of the thesis

In order to move closer to developing a new test paradigm for assessing hearing-impaired individuals' difficulty partaking in conversation, this thesis aimed to investigate whether there are objective measures of speech production and temporal dynamics, jointly referred to as conversational dynamics, that are changing systematically when people communicate in challenging conditions. In all our four studies, the following conversational dynamics were measured: the participants' articulation rates, their speech levels, their accuracy and precision of turn-timing, their rate of turn-takings, and the duration of their interpausal units (IPUs), i.e., units of connected speech surrounded by silence. Table 1.1 presents an overview of the four studies.

Chapter 2 presents our first study, where we established the baseline of the direction of conversational dynamics when young NH interlocutors were exposed to challenging conditions. The participants elicited dialogue by solving the DiapixUK task (Baker and Hazan, 2011), a spot-the-difference task where people have to work together to find differences between two almost identical pictures. We manipulated the degree of expected communication difficulty by having the participants communicating in their first and second language (Danish; L1 and English; L2, respectively) and in quiet and in the presence of speech-shaped background noise (ICRA7, Dreschler et al., 2001). The participants were separated into booths, communicating via headphones and microphones, and in noise conditions, they heard the noise through their headphones.

In *Chapter 3*, we made a smaller, exploratory study investigating the impacts of noise and conversational task on conversational dynamics between NH interlocutors conversing in their L2. The setup was similar to that in the first experiment. As the data was collected by a native speaker of English who was monitoring the conversations, interlocutors were only conducting the experiment in their L2. In half of the conditions, the interlocutors elicited dialogue by engaging in the DiapixUK task; in the other half, they spoke freely with the aid of some predefined conversational topics available to them if they needed inspiration. Half of the conversations were in quiet, the other half in the presence of 20-talker babble created from recordings from the first study.

In *Chapter 4*, we investigated the effects of increasing noise level and hearing impairment on conversational dynamics in a setup similar to the first two studies. Pairs consisting of a young NH individual and an older HI individual solved the DiapixDK task (Sørensen, 2021), a Danish version of the DiapixUK task, which we translated for this study. The interlocutors conversed either in quiet, or in the 20-talker babble used in the second experiment presented at 60, 65, and 70 dBA SPL.

In our last experiment presented in *Chapter 5*, we investigated the effects of noise and hearing aid amplification on conversational dynamics between pairs of young NH and older HI individuals. In this experiment, interlocutors sat face-to-face, solving the DiapixDK task. In half of the conditions they conversed in quiet; in the other half they conversed in the presence of the 20-talker babble. In half of the quiet and noise conditions, the HI participants received simple amplification via a hearing aid, and in the other half they received no compensation for their hearing loss.

Finally, a summary and discussion of the main findings of this thesis is presented in the concluding *Chapter 6* along with future directions and perspectives for using measures of conversational dynamics in the development of hearing assistive devices.

Chapter	**Participants**	**Conditions**	**Task**	**Noise type**	**Modality**
2	NH/NH	Noise Second language	DiapixUK	Speech-shaped noise	Audio-only
3	NH/NH	Noise	DiapixUK Free conversation	20-talker babble	Audio-only
4	NH/HI	Noise at different levels	DiapixDK	20-talker babble	Audio-only
5	NH/HI	Noise Amplification	DiapixDK	20-talker babble	Face-to-face

Table 1.1: Comparison of the four studies presented in this thesis.

2

The effects of noise and second language on conversational dynamics in task dialogue[a]

Abstract

This study provides a framework for measuring conversational dynamics between conversational partners (interlocutors). Conversations from 20 pairs of young, normal-hearing, native-Danish talkers were recorded when speaking in both quiet and noise (70 dBA SPL), and in Danish and English. Previous studies investigating the intervals from when one talker stops talking to when the next one starts, termed floor-transfer offsets (FTOs), suggest that typical turn-taking requires interlocutors to predict when the current talker will finish their turn. We hypothesized that adding noise and/or speaking in a second language (L2) would increase the communication difficulty and result in longer and more variable FTOs. The median and interquartile range (IQR) of FTOs increased slightly in noise, and in L2 there was a small increase in IQR, but a small decrease in the median of FTO durations. It took the participants longer to complete the task in both L2 and noise, indicating increased communication difficulty. The average duration of inter-pausal units, i.e. units of connected speech surrounded by silences of 180 ms or more, increased by 18% in noise and 8% in L2. These findings suggest that talkers held their turn for longer, allowing more time for speech understanding and planning. In L2, participants spoke slower, and in both L2 and noise, they took fewer turns. These

[a] This chapter is based on Sørensen, A. J. M., Fereczkowski, M, & MacDonald, E. N. *The effects of noise and second language on conversational dynamics in task dialogue* (in press). The data was collected as part of AJMS's M.Sc., but the analysis and interpretations were done as part of this PhD work.

changes in behavior may have offset some of the increased difficulty when communicating in noise or L2. We speculate that talkers prioritize the maintenance of turn-taking timing over other speech measures.

2.1 Introduction

Traditionally, hearing research involving speech has focused mainly on experiments where either speech perception or production is measured in isolation. However, conversation is a complex collaborative effort involving an overlap between comprehension and production, along with feedback and adaptation processes that occur both within and between interlocutors (i.e., conversational partners). These adaptations can include responses to the environment and each other's behaviour, such as the opportunity to repair errors by signalling difficulties in understanding (Schober and Clark, 1989; Wilson and Wilson, 2005). While the field of conversational analysis has investigated many aspects of interactive communication, it has traditionally focused on conversations conducted in favourable acoustic environments with normal-hearing (NH) interlocutors. However, recent studies have started to investigate how some factors, which are known to affect speech intelligibility, influence conversational behaviour (e.g., Aubanel et al., 2011; Beechey et al., 2018, 2020a,b; Hadley et al., 2019). The motivation for the present study was to investigate if more challenging communication conditions influenced the timing of turn taking in conversation.

The fundamental organization of a conversation is based on a structure where people take turns in an alternating fashion with each other. The timing in turn taking can be quantified by the floor-transfer offset (FTO), which is defined as the interval from when one person stops talking to when the next person starts talking. This interval can either be negative, indicating an acoustic overlap of the interlocutors' speech signals, or positive, indicating an acoustic gap between the speech signals. The FTO distribution from Levinson and Torreira (2015) can be seen in Figure 2.1 and is representative of the distributions that have been observed in other studies (e.g., Aubanel et al., 2011; Brady, 1968; Heldner and Edlund, 2010; Norwine and Murphy, 1938; Stivers et al., 2009). In general, these distributions are unimodal and right-skewed, with a peak around 200 ms.

In order to achieve the FTOs observed in these studies, it has been argued that talkers predict when interlocutors will end their turns, and this is supported

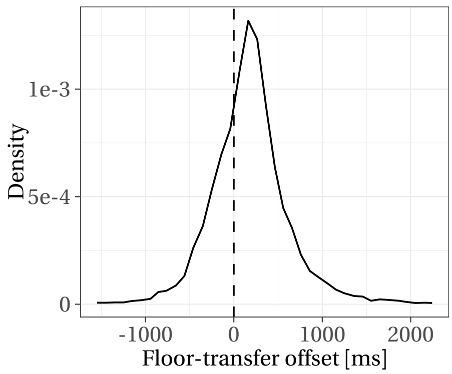

Figure 2.1: Distribution of floor-transfer offsets (FTOs) from about 38 hours of spontaneous dialogue in English from The Switchboard Corpus taken from Levinson and Torreira (2015). Data has been adapted to show the density instead of frequency of FTOs.

by the results from many different studies. First, the latency of speech production is larger than the modal response times observed in FTO distributions. Preparing to articulate a single word takes about 600 ms, and well over one second for multi-word utterances (Indefrey and Levelt, 2004; Magyari et al., 2014). When investigating the timing of in-breaths prior to answers, Torreira et al. (2015) observed that when preparing for a short response, participants answered on residual air, whereas for longer responses, they inhaled, and the average inbreath timing was 15 ms after the end of the questioner's utterance. They argue that since initiating inhalation takes 140-320 ms (Draper et al., 1960), this implies that the duration of the response was planned during an interlocutor's utterance. Bögels et al. (2015b) studied event-related brain potentials (ERPs) from EEG data during an interactive quiz. They manipulated the placement of the critical information for answering quiz questions either midway through or near the end of a sentence. In one condition participants had to respond to the questions, and in the other they only listened to the sentences. Compared to passive listening, when listeners had to respond, positivities in the ERPs at the point of the critical information (either early or late in the question) were observed in brain areas that are associated with language production. This suggests that people start planning their response as soon as they can. Further, Bögels et al. (2015b) found evidence of switches in attentional resources between comprehension and production in the conditions where the critical information was presented early in the question (i.e., when participants started planning their response in parallel with listening to the remainder of the question). When conducted as a divided attention task, Boiteau et al. (2014) found evidence of deteriorated visuomotor tracking-performance near the end of an interlocutor's turn or the start of one's own turn, corresponding to the points in conversation that are most cognitively demanding. For further review of the evidence of response planning and a model of comprehension and production during turn-taking, see Levinson and Torreira (2015).

Some studies have identified acoustic cues that are used to predict turn ends. De Ruiter et al. (2006) asked participants to press a button when they anticipated that a talker's turn would end when listening to excerpts from recorded conversations that had been processed in different ways. By comparing the prediction performance across the conditions, they demonstrated that both lexicosyntactic and prosodic cues are used to predict turn ends. Gravano and Hirschberg (2011) identified several acoustic cues that they associated with turn yielding. They compared utterances that led up to turn-switches with utterances leading up to turn holds. They found the following cues to predict turn-switches well: a point of textual completion (i.e., the point where an utterance can be grammatically complete), a reduction in intensity level, a reduction in pitch level, a falling or rising intonation at the end of an utterance, a reduced lengthening of the final words in an utterance, as well as increased vocal jitter, shimmer, and noise-to-harmonic ratio. They further found that the larger the number of these cues that were present in the utterance, the more likely it was to yield a turn-switch.

As outlined above, in order to respond rapidly and maintain fluid turn taking in conversation, listeners must simultaneously process the incoming acoustic signal to understand what is being said, plan a response, and predict when their interlocutors will end their turns. In the present study, we hypothesized that reducing processing resources by making conversation more challenging would alter turn-taking behaviour. In this study, we tested this by manipulating the degree of expected communication difficulty. Conversations were recorded both in the absence and presence of background noise, with talkers speaking both in their native language (L1; Danish), and in a second language (L2; English). Given that interlocutors have limited processing resources, we hypothesized that making conversation more challenging would alter the FTO distribution. For example, listening to speech in the presence of noise or in a second language may require increased listening effort, reducing the resources available to plan speech and predict turn ends. This could both delay the articulation of responses (shifting the FTO distribution to the right) and increase the variability in the timing of the floor transfers (broadening the FTO distribution). In isolation, while speaking in noise should not increase the difficulty of speech planning, speaking in L2 may be more difficult (e.g., García Lecumberri et al., 2017; Wester et al., 2014), resulting in longer and more variable FTOs. The ability to predict the timing of turn ends may also be reduced in noise or in L2. Previous studies have demonstrated that listeners use both lexicosyntactic and prosodic cues to predict the timing of turn ends (Brusco et al., 2020; De Ruiter et al., 2006; Gravano and Hirschberg, 2011; Riest et al., 2015). Compared to when listening in L1, processing the lexicosyntactic cues used to predict turn ends may be more difficult in L2. Depending on how similar they are between languages, the saliency of the prosodic cues may or may not differ between L1 and L2 (e.g., Brusco et al., 2020). Thus, the impact of L2 on predicting the timing of turn ends may vary between languages. Compared to when listening in quiet, listening in noise may reduce the saliency of both lexicosyntactic and prosodic cues. Increased variability in predicted timing of turn ends, due to noise or conversing in L2, could lead to more variable timing of floor transfers (i.e., a broader FTO distribution).

2.2 Methods

2.2.1 Participants

In this study, forty normal-hearing native-Danish talkers ($\mu = 26$ years, $\sigma = 3.7$ years, 12 women) participated in pairs (four mixed-gender pairs). Participants within each pair knew each other well. Standard audiograms were measured for all participants ensuring their hearing threshold levels were below 20 dB HL between 125 Hz and 8 kHz. All participants reported being "comfortable" in English and had all participated in at least one university-level class taught in English. All participants provided informed consent and the experiment was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). The participants were compensated for their time.

2.2.2 Conversational task and conditions

Dialogue was elicited by conducting the DiapixUK task (Baker and Hazan, 2011), a spot-the-difference task in which pairs are given almost identical cartoon pictures, and they have to work together to find the differences between them. Using this task provides several advantages. First, the completion time can be measured. Second, the content is more limited than free conversations, making the conversational content more homogenous across pairs. Finally, the task requires both talkers to communicate, potentially leading to both talkers speaking more equally compared to free conversation, where one talker might dominate the conversation.

In total, the participants conducted the DiapixUK task in four different conditions: in L1 and L2, both in quiet and in a noise background, consisting of a 6-talker speech-shaped noise (ICRA 7, Dreschler et al., 2001).

2.2.3 Setup

The talkers sat in two separate sound-treated booths and communicated over headphones and microphones. The talkers were unable to see each other during the experiment. Thus, the participants only had access to acoustic turn-taking cues. Each participant wore Sennheiser HD650 open headphones and Shure WH20 headset microphones placed close to the mouth at the position recommended by the manufacturer.

Recordings

An operator sat outside the booths monitoring the experiment and could communicate with the participants through an operator microphone. In the headphones, the participants heard a mix of 1) themselves, 2) their interlocutor, 3) the operator (only if the operator needed to talk), and 4) the background noise (only in the noise conditions). The signals from this headphone mix, the individual Shure microphones, and the operator microphone were recorded on four separate channels using an RME Fireface 802 soundcard. Each signal was sampled at 48 kHz with a bit depth of 24, using MATLAB 2016a. All the recordings for which we have received consent have been made publicly available (Sørensen et al., 2018).

Calibration

The noise was calibrated to an average presentation level of 70 dBA SPL in the headphones. The level was calibrated by placing the headphones on a B&K 4149 microphone preamplified by a B&K 2619 (hereafter called the headphone coupler) connected to a B&K 2636 sound level meter (SLM). As the level of the ICRA7 noise fluctuates continuously, a 10 s integration time was used in the SLM to obtain an overall presentation level of 70 dBA SPL.

The levels of both microphones were calibrated such that the broadband, A-weighted levels presented over the headphones were the same as if their interlocutor was one meter away from them in the same room. To do this, a Nor140 SLM was placed one meter from a talker and the headphones were placed on the headphone coupler (connected to the B&K SLM). While a talker produced a prolonged vowel, the gain from the headset microphone was adjusted in RME Totalmix such that the A-weighted levels measured from both SLMs were equal.

2.2.4 Procedure

Prior to the test session, a two-step training session was conducted. First, to familiarize participants with the task, they conducted a Diapix task using pictures from the original Diapix corpus (Van Engen et al., 2010) while facing each other outside the audiometric booths and under the operator's supervision. Following this, they moved to the two separate booths and conducted a second Diapix task, again using different pictures from the original Diapix corpus. During this part of the training, background noise was added to the communication channel.

The test session consisted of three blocks (repetitions) of four conditions consisting of the combinations of conversing in either their first (L1; Danish) or second language (L2; English) in quiet or noise. The order of the conditions was randomized within each block. After each block, the participants had a break. For each condition in each block, the pairs looked for 10 differences in a pair of DiapixUK pictures. The participants were not instructed in any particular strategy or encouraged to solve the task as quickly as possible. They were only instructed that they had 10 minutes to find 10 differences and that the experiment would proceed to the next condition if not completed within that time frame. However, all pairs were able to complete the task in less than 10 minutes in every condition. The 12 image pairs of the DiapixUK were counterbalanced across conditions and pairs. Thus, over the entire experiment, each DiapixUK image pair appeared five times in each condition, and 12 conversations were recorded from each participant pair.

2.2.5 Analysis of recordings

Each wave file was processed to automatically categorize and label the speech segments of each talker into different conversational categories following variations of the algorithms used by Heldner and Edlund (2010) and Levinson and Torreira (2015). An illustration of the categorization of conversational states can be seen in Figure 2.2.

First, the individual microphone tracks for each talker were processed to determine when each person spoke using Voice Activity Detection (VAD). The speech streams were buffered into segments of 5 ms with 1 ms overlap, and the RMS in each segment was computed. Based on a threshold value, segments were either labelled with 1 (speech) or 0 (no speech). The threshold value was determined individually for each talker in each conversation by hand. Following the procedure in Heldner and Edlund (2010), gaps smaller than 180 ms were bridged in order to minimize the risk of mistaking stop consonants for pauses between speech units. Any sound bursts shorter than 70 ms were set to 0 as they were assumed to be non-speech (e.g. coughs).

Next, for each conversation, the binary speech/no-speech streams from the two talkers were fed into a conversational state classification algorithm developed for this study. The algorithm labelled speech into the following categories: gaps (joint silences of both talkers during a floor transfer), overlaps-between (overlapping speech during a floor transfer), overlaps-within (speech where the utterance of one talker is completely overlapped by speech from the other and there is no floor transfer), pauses (joint silence not followed by a floor transfer), interpausal units (IPUs; units of connected speech in which any included acoustic silences are less than 180 ms), and turns (sequences of IPUs by one talker surrounded by floor-transfers). The FTO distributions were measured, along with the rate at which floor-transfers occurred. Moreover, the rate at which overlaps-within occurred was measured. In order to verify that the state classification algorithm worked as intended, one conversation was manually labeled with the categories presented above (and in Figure 2.2) and compared to the automated analysis to make sure they agreed. Further, approximately 60% of all the overlaps-within (3210 out of the total 5171) across the four conditions were manually annotated to investigate differences in those overlaps across conditions. In this process, it was confirmed that the algorithm had labeled the overlaps-within correctly.

To estimate the speech levels for each talker, the following procedure was used. First, the RMS of the headphone mix of noise-only segments in conversations carried out in background noise was calculated. Since the noise was presented at 70 dBA SPL, this RMS was used to calculate a conversion factor from dB FS to dBA SPL for the headphone mix wav file. Next, for each talker, the recordings of conversations in quiet were examined to identify segments where only speech from that talker was present in the headphone mix. For these segments, the RMS from the headphone mix was compared to that from the talker's close mic. Based on this and the previously calculated conversion factor for the headphone mix, a conversion factor from dB FS to dBA SPL was calculated for the talker's close mic. Finally, the speech level for each individual was calculated by measuring the RMS recorded by the close mic for all the speech units excluding pauses, and the conversion factor was used to convert the RMS to estimated speech levels in dBA SPL. In all conversations, the number of syllables produced by the individual talkers were computed using the Praat (Boersma and Weenink, 2017) script presented in De Jong and Wempe (2009) with default parameter settings. The algorithm detects syllable nuclei (the peak within the syllable) using measures of intensity and voicedness. It extracts the intensity and considers only peaks above a threshold corresponding to the median intensity over the whole sound file. Of these peaks, only the peaks that have a preceding dip of at least 2 dB with respect to the current peak are considered. Finally, to exclude voiceless consonants, the syllable nuclei are extracted by excluding unvoiced peaks found by the pitch contour. Using an interface between Praat and MATLAB 2020b, the Praat-detected syllable nuclei were extracted from Praat TextGrids for each person in each of their conversations. To estimate the articulation rate of each talker, the number of syllables identified using the Praat script was divided by the phonation time determined by the VAD described above.

2.2.6 Statistical procedure

For analyzing the effects of noise, second language and replicate on various measures, linear mixed-effects regression models were fitted to the variables of interest using the *lme4* package in *R* (Bates et al., 2015). Unless otherwise stated, the starting model consisted of background (quiet, noise), language (L1, L2), and replicate (1, 2, 3) as fixed effects with up to third order interaction, and a random intercept varying among pair and person within pair, i.e. the starting model was:

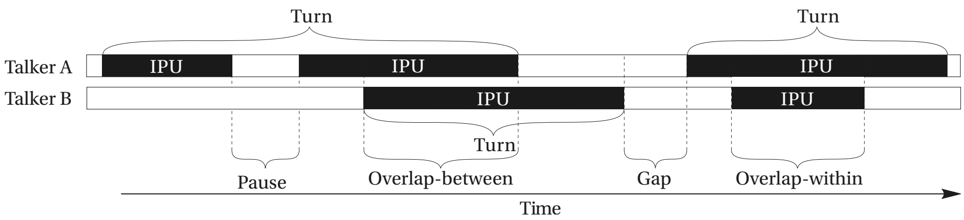

Figure 2.2: Illustration of the classification of gaps, overlaps-within, overlaps-between, pauses, interpausal units (IPUs) and turns during conversations between Talker A and B. A person's turn is measured from the onset of the IPU following a floor-transfer to the offset of the IPU followed by a floor-transfer. There are two floor-transfer offsets (FTOs): the overlap-between and gap.

$x \sim$ background $\times$ language $\times$ replicate $+ (1 \mid$ pair/person). The *interaction.plot* function from the stats package was used to judge whether random slopes for any of the predictors should be included in the starting model. The *lmerTest* package in *R* (Kuznetsova et al., 2017) was used to perform backward elimination of both fixed and random effects of the models. This was done by first defining the largest model as described above and using the step function to reduce the model by first simplifying the random-effects structure and afterward the fixed-effects structure in a step-wise manner by deleting model terms with high p-values. The *anova* function from the stats package in *R* as well as residuals plots were used to compare models before and after reducing them to find the model that best fit the data. Finally, ANOVA tables were computed with Satterthwaite approximated denominator degrees-of-freedom (df) corrected F-tests for the fixed effects. The *lsmeans* function from the *lmerTest* package was used to compute pairwise comparisons of least-squares means of the significant effects using the Satterthwaite approximated df.

2.3 Results

2.3.1 Speech production and task completion time

The levels of speech one meter from the talkers were estimated from the individual recordings, and are plotted in Figure 2.3. The final selected model describing the speech level was as follows: speech level $\sim$ background + replicate + (1 | pair/person). The speech level increased by an average of 9.4 dB in noise [$F(1{,}437) = 6061$, $p < .001$], but there was no effect of L2 on the speech level (the language factor was eliminated from the full model by the *step* function [$F(1{,}436) = 2.27$, $p = .132$]. There was a significant effect of replicate [$F(2{,}437)$

= 4.76, $p < .01$]. A post-hoc analysis showed that this effect was driven by a significant decrease of 0.46 dB between replicates 1 and 3 [$t(437) = 3.07$, $p < .01$]. The average speaking level in noise was 67.5 dBA SPL, resulting in an average SNR of -2.5 dB.

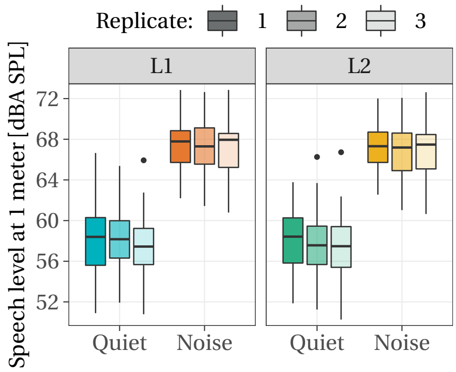

Figure 2.3: Boxplots of average speech levels one meter away from the talkers (in dBA SPL). The results are presented as averages across participants in the three replicates of the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2. Here and in later plots, the boxplots show the 25th, 50th (median), and 75th percentile, and the whiskers indicate minimum and maximum observations. Outliers are observations above or below 1.5 times the interquartile range.

To estimate the reliability of the Praat script developed by De Jong and Wempe (2009) that was used to calculate the articulation rates, the number of syllables in three IPUs per replicate for all participants in all conditions were manually counted (MC) and compared to the number of syllables computed by the Praat script (PC). As a selection criterion, a person's first three utterances in a conversation that Praat had detected to have a minimum of 0 and a maximum of 11 syllables were selected for the analysis. The lower bound was chosen because the seldom occurring ingressive "ja" ("yes") in Danish does not have a syllable nucleus and will not be detected by the script as a syllable. The upper bound made it easier for the listener to maintain the syllables in memory when manually counting them. For L1 in quiet, L1 in noise, L2 in quiet, and L2 in noise, the average PC/MC ratios were 0.97, 0.97, 0.98, 0.98 (with standard deviations of 0.077, 0.067, 0.079, 0.081), respectively. A linear mixed-effects model showed that there was no effect of language [$F(1,475) = 2.03$, $p = .155$], background [$F(1,475) = 0.06$, $p = .81$], or interaction between the two factors: [$F(1,475) = 0.04$, $p = .837$], and no effect of person: [$F(1,475) = 1.52$, $p = .218$] on the ratios.

Boxplots of the articulation rates are depicted in Figure 2.4. The final selected model was as follows: articulation rate ∼ language + (1 + background + replicate | pair) + (1 + language | pair/person). An average decrease in articulation rate by 0.5 syllables/second in L2 compared to L1 was statistically significant [$F(1,39) = 302$, $p < .001$].

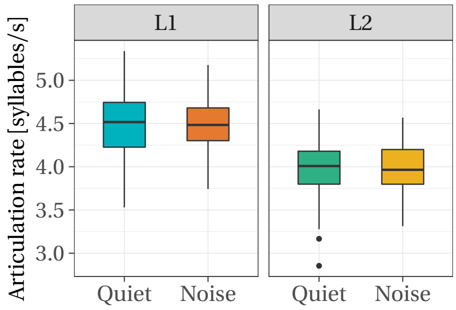

Figure 2.4: Boxplots of articulation rates of the talkers measured in syllables/second in the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2. Syllable nuclei were detected per person using the Praat script presented in De Jong and Wempe (2009) and were divided by the person's phonation time.

The task-completion time, i.e., the time it took each pair to find 10 differences between the Diapix, was measured. Figure 2.5 shows boxplots of the completion time in the four conditions and three replicates. A random intercept for the Diapix picture pairs was added to the starting model as the difficulty of the task could vary across Diapix picture pairs. The final model was as follows: completion time ∼ background + language + replicate + (1 + replicate | pair) + (1 | picture). There was a statistically significant training effect, i.e., the average completion time decreased with replicate [$F(2,29.6) = 12.6$, $p < .001$]. A pairwise comparison post-hoc analysis revealed a significant difference between the first and second replicate [$t(20.5) = 3.91$, $p < .001$], and between the first and third replicate [$t(18.9) = 5.03$, $p < .001$], but only a borderline significant decrease between second and third replicate [$t(68.6) = 1.96$, $p = .054$]. During the experiment, the operator observed that over the course of the first block, pairs discovered that the primary differences between images often involved signs or colors. As a result, they changed the order in which they searched the images and became quicker at solving the task. The completion time in noise compared to quiet increased significantly by, on average, 31 seconds [$F(1,186) = 16.8$, $p < .001$]. Similarly, the task completion time increased significantly by, on average, of 47 seconds in L2 compared to L1 [$F(1,186) = 39.9$, $p < .001$].

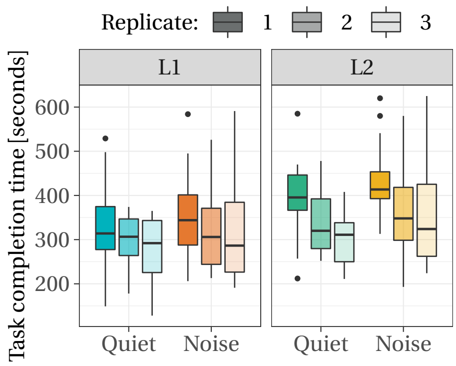

Figure 2.5: Boxplots of the time it took the pairs to complete the task in the three replicates of the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2.

In summary, the participants spoke louder and took longer to complete the task in noise. When comparing L2 to L1, they spoke at the same level, but slower and took longer to complete the task. Finally, they completed the task faster in the second and third replicate compared to the first, and they spoke slightly softer in the third replicate.

2.3.2 Floor-transfer offsets

The overall hypothesis was that with increased processing demands, we would see a delay and more variability in the timing of people's turn-taking. As a measure of centrality of the distribution, the median was used rather than the mean as FTO distributions are slightly positively skewed. For the same reason, the interquartile range (IQR) was used rather than the standard deviation as a measure of variability.

Kernel density plots (computed using *geom_density* from the *ggplot2 R* package) were computed for the FTOs in each condition pooled across all pairs (see Figure 2.6). Descriptive statistics of the distributions are provided in Table 2.1. As seen in Figure 2.6, the pooled distributions look highly similar. The final selected model for analyzing the median of FTOs was as follows: median FTO $\sim$ background + language + $\left(1 + \text{background} \mid \text{pair}\right)$. There was a borderline significant increase of 21 ms in noise: [$F(1,19) = 4$, $p = .06$], and a significant decrease of 19 ms in L2: [$F(1,439) = 9$, $p < .01$]. We observed that the proportion of overlaps during turn-taking increased in L2 [$F(1,81.4) = 38.5$, $p < .001$], contributing to the observed decrease in median FTO in L2.

Boxplots of the IQR are plotted in Figure 2.7, right panel. The analysis was done on log-transformed IQRs to meet the residual-normality assumption of the linear model. The final model for analyzing the IQR of FTOs was the following: log(IQR FTO) $\sim$ background + language + $\left(1 + \text{background} \mid \text{pair/person}\right)$. There were significant increases in IQR in L2, 14 ms [$F(1, 399) = 4.5$, $p < .05$], and in noise, 41 ms [$F(1, 39) = 15.6$, $p < .001$]. We computed the floor-transfer rate on a pair-level, see Figure 2.8. The final selected model describing the number of floor-transfers per minute was the following: FT rate $\sim$ background + language + replicate $\left(1 + \text{background} + \text{language} + \text{replicate} \mid \text{pair}\right)$. The rate of floor-transfers per minute decreased by 1.6 in noise [$F(1, 19) = 18.3$, $p < .001$], and by 2.6 in L2 [$F(1, 23) = 57.5$, $p < .001$], and increased with replicate [$F(2, 20.5) = 4{,}6$, $p < .05$]. There was a significant increase between replicates 1 and 3 of 1.3 occurrences/minute [$t(19.5) = -2.8$, $p < .05$], and of 1 occurrence/minute between replicates 2 and 3 [$t(22.3) = -2.31$, $p < .05$].

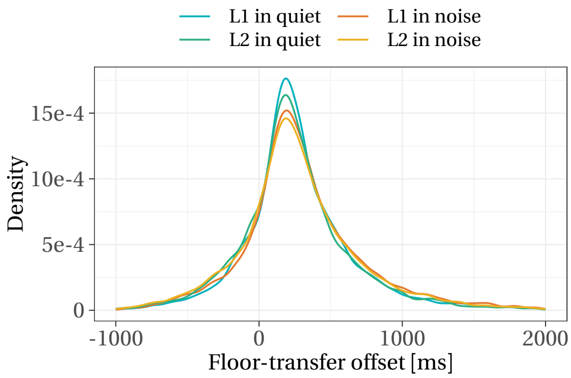

Figure 2.6: Kernel density plots of the floor-transfer offsets (FTOs) pooled across pairs and replicates in the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2. Negative FTOs indicate acoustic overlap of the two talkers while positive FTOs indicate acoustic gaps.

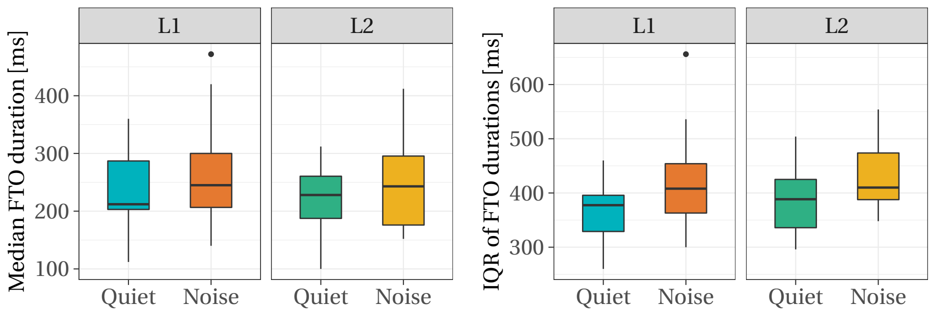

Figure 2.7: Boxplots of median floor-transfer offset (FTO) (left panel) and interquartile range of FTOs (right panel) in each of the four conditions: first language (L1) and language (L2) in quiet and noise.

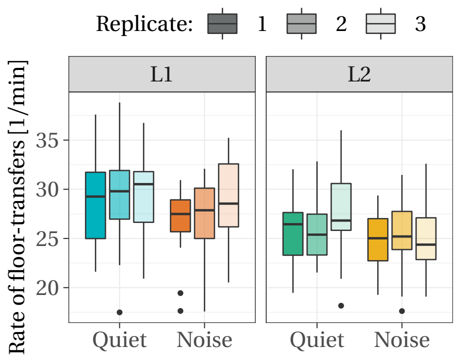

Figure 2.8: Boxplots of the number of floor-transfers per minute in the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2.

Condition	*N*	% overlap-between	% gap	Mean	Median	Mode*	Skewness	IQR	Min	Max
L1 in quiet	9036	17.4	82.6	298	232	190	2.30	364	-2384	7136
L1 in noise	9143	17.5	82.5	332	252	192	1.95	424	-2248	5920
L2 in quiet	9471	20.7	79.3	262	220	187	1.44	384	-1980	4868
L2 in noise	9588	20.6	79.4	294	240	191	1.66	436	-2092	6228

Table 2.1: Descriptive statistics of the floor-transfer offsets (FTOs) in ms in the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2. *Modes are calculated by taking the max of Gaussian kernels computed using the *density* function in *R*.

2.3.3 Interpausal unit durations

The final model for analyzing the duration of IPUs was as follows: median IPU $\sim$ background + language + $(1 + \text{background} + \text{language} + \text{replicate} \mid \text{pair/person})$.

There was a statistically significant increase of the median IPU duration of about 18% in noise [$F(1, 50) = 86.7$, $p < .001$], and of about 8% in L2: [$F(1, 40) = 15.4$, $p < .001$]. The increase in median duration appears to be driven by a general lengthening of all IPUs, rather than just a reduction in the frequency of very short (e.g., one syllable) IPUs. This is indicated by a shallower slope of the pooled IPU durations across pairs as seen in Figure 2.9, left panel, where the density has been log-transformed to more easily compare the distributions.

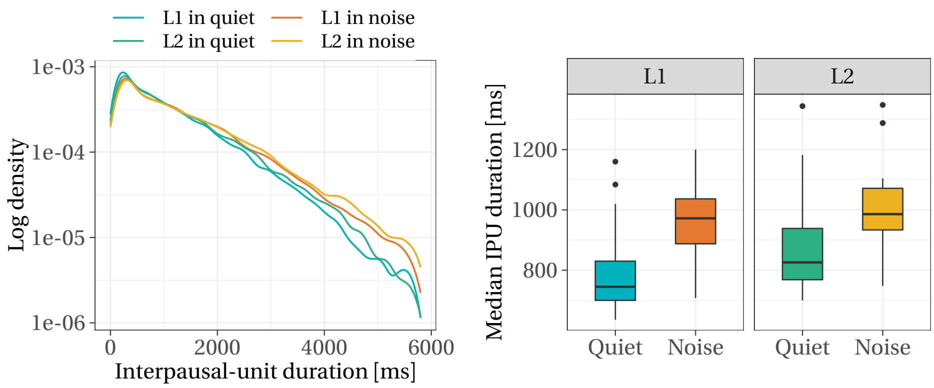

Figure 2.9: Kernel density plots with a logarithmic y-axis of the interpausal units (IPUs) pooled across pairs and replicates (left panel), and boxplots of median durations of IPUs (right panel) in the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2.

2.3.4 Overlaps-within

The duration of overlaps-within, i.e. utterances from talkers that temporally occur completely within utterances of their interlocutors (see Figure 2.2), had a small but significant increase of 15 ms in L2 [$F(1, 438) = 5.4$, $p < .05$], see Figure 2.10, left panel. The analysis was performed on log-transformed overlaps-within durations, and the final model was as follows: log(median OW) $\sim$ language + (1 | pair/person). We computed the rate of overlaps-within for each person in each conversation as the sum of occurrences of overlaps-within divided by the total duration of that conversation. The final selected model was as follows: OW rate $\sim$ language + $\left(1 + \text{background} \mid \text{pair}\right)$. The increase in the rate of overlaps-within by about 0.3 occurrences/minute in L2 was significant: [$F(1, 439) = 17.2$, $p < .001$], see Figure 2.10, right panel.

To further investigate possible differences across the four conditions, we listened to a subset of the overlaps-within (3210 out the total of 5171) across the four conditions and annotated them using a combination of the categories introduced in Gravano (2010), Levinson and Torreira (2015), and Schegloff (2000). The following categories were used and are not mutually exclusive (e.g., an overlap-within could feature an attempt to take a turn and also exhibit repeated syllables/words):

1. **Simultaneous start:** overlap occurs within 200 ms from the onset of the overlapped talker's utterance
2. **Verbal backchannels or agreements:** e.g. "yeah", "right", "uh-huh"
3. **Non-verbal backchannel:** e.g. laughter
4. **Continuation:** the overlapping interlocutor continues his previously acoustically terminated turn while the other talker took the floor before the onset of the overlap-within
5. **Attempt to take the turn**
6. **Incomplete turns:** turns ending mid-word/mid-utterance
7. **Repeated syllables/words:** repetition of words or syllables during or close to the overlap interval

In Table 2.2 the pooled frequency of the different overlap-within features can be found. Mixed-effects models with language and background as fixed effects with interaction and pair as random intercept were fitted to six of the seven features. Again, the *step* function in *R* was used to reduce the models, and an ANOVA analysis was performed with Satterthwaite approximated denominator degrees-of-freedom (df) corrected F-tests for the fixed effects.

There was no difference between conditions in the frequency of verbal backchannels. As the occurrence of non-verbal backchannels was very rare, there were many conversations in which this did not occur, and therefore no statistical test was performed. In noise, there were significantly more attempts to take the turn [$F(1,58) = 25.8$, $p < .001$], incomplete turns [$F(1,59) = 20.6$, $p < .001$], and repeated words/syllables [$F(1,58) = 7.9$, $p < .01$]. In L2, a decrease in the rate of simultaneous starts bordered significance: [$F(1,58) = 3.86$, $p = .054$], and there were significantly more repeated words/syllables [$F(1,58) = 14$, $p < .001$]. There was a significant interaction between background and language on the frequency of continuations [$F(1,57) = 6.93$, $p < .01$]. A post-hoc analysis

	L1 in quiet	L1 in noise	L2 in quiet	L2 in noise
Simultaneous starts	34.1 %	32.5 %	31.9 %	29.2 %
Verbal backchannels	55.6 %	55.3 %	57.5 %	55.4 %
Non-verbal backchannels	3.1 %	1.9 %	4.2 %	1.9 %
Continuations	12.1 %	7.9 %	7.8 %	8.2 %
Attempts to take the turn	29.3 %	36.8 %	27.6 %	35.9 %
Incomplete turns	19.5 %	27.6 %	20.9 %	28.4 %
Repeated words/syllables	3.2 %	8.6 %	6.8 %	10.4 %

Table 2.2: Frequency of overlap-within features, hand labeled from judging 3695 overlaps-within ($N_{\text{L1 in quiet}} = 801$, $N_{\text{L2 in quiet}} = 801$, $N_{\text{L1 in noise}} = 803$, $N_{\text{L2 in noise}} = 805$).

showed there were fewer continuations in noise in L1 than in quiet in L1, and significantly fewer continuations in quiet in L2 than in L1.

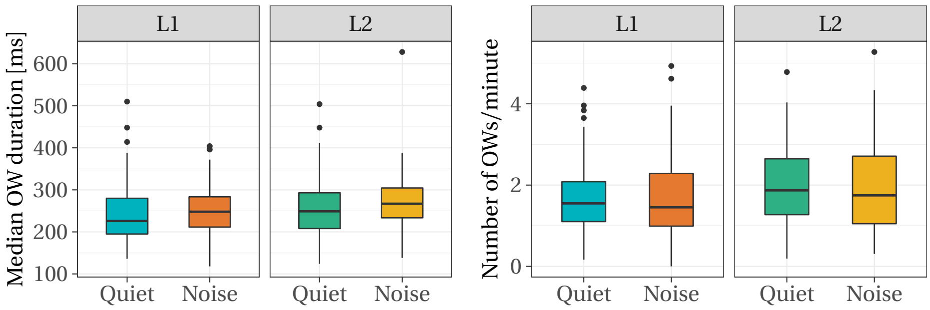

Figure 2.10: Boxplots of durations of overlaps-within (OW, left panel) and number of OWs per minute (right panel) in the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2.

2.3.5 Floor-transfer duration vs. interpausal unit duration

In Figure 2.11, the distribution of FTOs for different quartiles of IPU durations preceding floor-transfers (left panel) and following floor-transfers (right panel) are plotted. For the IPUs preceding floor-transfers, the tendency was that the longer their duration, the shorter and less variable the FTO was, indicated by the narrower distribution and slightly shifted peak. For the IPUs following floor-transfers, we see the opposite: the shorter in duration they were, the shorter and less variable the FTO, indicated by the narrower distribution and slightly shifted peak. A statistical analysis of the changes in median and IQR of FTOs for the four quartiles of IPU durations before and after floor-transfers in the four conditions confirmed this trend (see Figures A.1 and A.2 as well as Table A.1 in Appendix A.1.2).

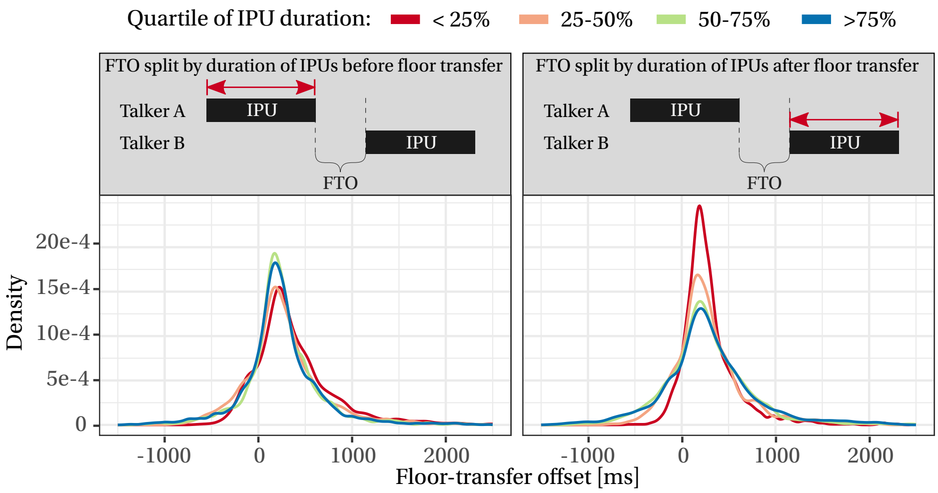

Figure 2.11: Distribution of FTOs for the four quartiles of preceding IPU duration (left panel) and following IPU duration (right panel) across all conditions.

2.4 Discussion

The goal of this study was to investigate how turn-taking behaviour changes when communication becomes more challenging. Here, two manipulations were used to increase the difficulty of communication: the presence of a background noise and conversing in L2. Talker pairs were asked to find ten differences between pairs of almost identical pictures. As expected, participants took longer to complete the task both in the presence of noise and when talking in their second language, indicating that both manipulations increased the difficulty of communication. While both noise and L2 influenced several other aspects of communication behaviour, the effects differed for some of the measures between the two manipulations.

In Figure 2.2, we illustrated the temporal dynamics between interlocutors in a dialogue. We presented data from three of these temporal dynamics in this study: interpausal units (IPUs), floor-transfer offsets (FTOs), and overlaps-within. We define IPUs as units of connected speech by the same person surrounded by silence of min. 180 ms. FTOs are durations of turn-takings measured from when the first person stopped talking to the next person started. Overlaps-within are talkspurts from one person that temporally occur completely within IPUs of their interlocutor and, thus, do not involve a floor transfer. Compared to when they spoke in quiet, talkers in noise increased their speech level, and produced longer IPUs. The FTO distributions of conversations in noise were slightly broader and the peak was slightly shifted to the right, with medians that were approximately 21 ms longer. The rate at which floor-transfers occurred decreased in noise. There was no change in the rate at which overlaps-within occurred in noise, but they consisted of more attempts to take a turn, more incomplete turns, and more repeated syllables. In L1, there were fewer continuations in noise than in quiet. The frequency of verbal backchanneling in these overlaps-within did not increase compared to the conversations in quiet.

Compared to when they spoke in their native language, talkers in L2 spoke slower, floor transfers occurred at a reduced rate, and they produced longer IPUs. The FTO distributions in L2 were slightly broader, and were shifted slightly to the left, with medians that were approximately 19 ms shorter. The rate as well as the duration of overlaps-within increased in L2. For these overlaps-within, there were fewer simultaneous starts, and the frequency of utterances with repeated syllables increased. In quiet, there were fewer continuations in L2 compared to L1.

2.4.1 Timing of turn-taking

While conversing listeners must simultaneously process the incoming acoustic signal to understand what is being said, plan a response, and predict when their interlocutors will end their turns. Given that interlocutors have limited processing resources, we hypothesized that making conversation more challenging would alter turn-taking behaviour, which could be observed in changes in the FTO distribution. For example, increased listening effort could reduce the resources available to plan speech and predict turn ends, shifting the FTO distribution to the right and/or broadening the FTO distribution. Further, the perceptual saliency of the acoustic cues used to predict the timing of turn ends may be affected by both noise and L2. This could also increase the variability in the timing of floor transfers, resulting in a broader FTO distribution.

We observed slightly longer FTOs of 21 ms in noise compared to quiet, and the IQR of FTOs increased by 41 ms in noise. In L2, the IQR increased slightly by 14 ms, However, the median of FTOs decreased by 19 ms in L2, opposite to what we hypothesised. While three of the four observations were in the hypothesised direction, the effects were relatively small.

While the present study was designed to investigate potential changes in the distribution of FTOs in response to the presence of noise or conversing in a second language, other factors can be considered. Roberts et al. (2015) investigated factors that influenced FTO duration and found that FTOs were shorter when replies were shorter, and for interactions that involve a response action, the FTO was shorter in replies to utterances that were longer in duration. Both of these observations suggest FTOs are affected by planning and/or listening effort. For the first observation, shorter replies should require less motor planning and lead to shorter FTOs. The second observation, that longer utterances provide more time to complete speech and response planning, is somewhat counter-intuitive. It has been estimated that both speech understanding and speech planning are 3-4 times faster than the articulation of speech (Wheeldon and Levelt, 1995). Thus, the rate of speech communication is limited by the rate of articulation, and not planning or understanding. As a result, a longer IPU from one's interlocutor can provide more time to complete speech understanding and response planning as those processes are faster than the rate of articulation produced by the interlocutor. Based on this, replies in response to long IPUs should occur quicker (i.e., with shorter FTOs) than those in response to short IPUs, as observed by Roberts et al. (2015), and we observed longer IPUs both in noise and L2. For the results in the present study, we conducted a similar investigation by comparing FTOs before and after IPUs of different durations. As was seen in Figure 2.11, the overall pattern was that when talkers produced longer IPUs, their interlocutors were more "on-time" with their responses and there was less variability in this timing, likely because they had more time to plan their response. Further, for IPU durations after floor-transfers, shorter "response" IPUs resulted in shorter and less variable FTOs, likely because shorter IPUs require less planning. Thus, this analysis suggests FTO distributions are influenced by the processing demands of speech understanding and speech planning.

A potential explanation for why we did not observe large changes in FTO distributions in the present study was that the more difficult conditions were not sufficiently challenging. On average, it took participants longer to complete the spot-the-difference task in the more challenging conditions (approximately 10% longer in noise, 15% longer in L2, and 25% longer in both), which suggests that these manipulations did, indeed, have an effect on communication. In noise, participants spoke significantly louder, and in L2 they spoke significantly slower. While the manipulations increased the completion time and changed aspects of their speech production, they may not have been challenging enough to see large delays or increases in the spread in turn-timing. Aubanel et al. (2011) found that when interlocutors conversed in the presence of a background conversational pair, they delayed their responses. They further observed both increased speech levels as well as decreased speech rates in the presence of a background pair. Moreover, in conversations between NH and HI interlocutors, Sørensen et al. (2021, submitted, see Chapter 4) found that both participant groups had delayed and more variable responses, as well as decreased speech rates and increased speaking levels in the presence of background noise. Beechey et al. (2018) concluded that acoustic-phonetic speech production changes are most sensitive to low to moderate degrees of communication effort, rather than higher-level, turn-taking behavior. Thus, it is possible that while the challenges faced in this experiment were sufficient to alter other aspects of speech communication, they may not have been sufficient to observe large changes in turn-timing between interlocutors.

However, a second possibility for why we did not observe the larger hypothesized changes in the FTO distributions is that some of the other changes in speech production and conversational behaviour that was observed in the more challenging conditions may have reduced processing demands. For example, when speaking in L2, talkers spoke slower, produced longer IPUs, and floor transfers occurred at a slower rate compared to L1. Wester et al. (2014) and García Lecumberri et al. (2017), also found that speakers of L2, when solving the Diapix task, adopted more hesitant speech with a lower proportion of speech turns, a slower speech rate, more elongations as well as more pauses. When speaking in noise, talkers increased the length of their IPUs and floor transfers occurred at a slower rate compared to in quiet. As described above, since the rate of speech articulation is slower than the rates of speech understanding and planning, longer IPUs can reduce processing load. Further, the rate at which floor transfers occurred was lower in noise than in quiet. Overall, these adaptations could reduce processing load for both the talker and the listener, allowing the talkers to achieve turn-taking timing that is more similar to that achieved in quiet in their first language.

Bögels et al. (2015b) found that the participants in their quiz game started the planning of their responses as soon as they received the critical information.

However, even if talkers may be able to reply well before their interlocutors have finished their turns, they may wait to take their turn at a specific, more socially appropriate time. It has been shown that people are sensitive to the timing of turn-taking, and FTO distributions are similar across languages and cultures (Kendrick and Torreira, 2015; Levinson and Torreira, 2015; Stivers et al., 2009). For example, Kendrick and Torreira (2015) and Roberts et al. (2011) found that small increases in gap length are more associated with negative/dispreferred responses. Thus, we speculate that maintaining the timing of turn-taking behaviour is important to achieve socially appropriate interactions, and when faced with more challenging communication conditions, interlocutors modify other aspects of their speech production and interaction that may help them to maintain the timing of their turn-taking behaviour.

2.4.2 Interpausal unit durations

In the present study, we define IPUs as connected portions of speech that are separated by acoustic silences with durations of at least 180 ms. In noise and L2, talkers increased the duration of their IPUs, with the median length increasing by approximately 18% in noise and 8% in L2. Further, the observed increase in median duration appears to be driven by a general lengthening of all IPUs, rather than just a reduction in the frequency of very short (e.g., one syllable) IPUs.

The increase in IPU duration may be due to an increase in filler words. Clark and Fox Tree (2002) outline how talkers predict upcoming delays in their speech planning and prepare to insert filler words, such as “uh” for short delays or “um” for longer delays, as well as prolonging syllables to signal they are continuing an ongoing delay. These filler words occur both at the phrase boundary and mid-utterance. When produced mid-utterance, talkers plan when to insert the filler words into their sentences. They argue that a talker can use filler words mid-sentence to signal that they want to keep the floor despite a delay in speech planning. We speculate that in challenging conditions where speech planning may be delayed, interlocutors can also make use of filler words at the start of their turn to achieve socially acceptable timing of floor transfers.

Other studies of conversations have observed that in noisier and more complex acoustic environments, talkers increase IPU durations, consistent with a strategy of “holding the floor” (Beechey et al., 2018, as well as the studies presented in Chapters 3-5). As was found above, holding the floor for longer may consequently provide both the talker and the listener with more time to prepare their responses. Beechey et al. (2018) argued that a “holding-the-floor” strategy by increasing utterance durations and speaking faster (which was observed in their study) may ease communication for the individual, because it reduces the need for the talker to listen in adverse environments. In contrast, Hadley et al. (2019) found that interlocutors shortened their utterances with increasing noise level. They argue that the differences observed between their results and other studies may be due to task differences. In Hadley et al. (2019) interlocutors held “free” conversations based on predefined topics, whereas in Beechey et al. (2018), Sørensen et al. (2020b), and the present study, interlocutors collaboratively solved a puzzle. However, Watson et al. (2020, see Chapter 3) observed an increase in IPU duration in noise both when interlocutors held free conversations and solved the DiapixUK task together. Unlike the other studies where background levels were held relatively constant over the entire course of a conversation, in Hadley et al. (2019), the average background level changed randomly between 54, 60, 66, 72, and 78 dB SPL every 15-25 s. One could speculate that while talkers may adopt a “holding-the-floor” strategy in relatively constant background noise levels, they may adopt a strategy that is more flexible when communicating in more variable background noise (e.g. Aubanel and Cooke, 2013; Aubanel et al., 2012). However, we note that in Hadley et al. (2019) acoustic pauses in speech streams that were shorter than 1.25 s were bridged. Thus, they defined utterances as portions of speech separated by pauses of at least 1.25 s. In contrast, the present study and some others (Heldner and Edlund, 2010; Watson et al., 2020) used an acoustic pause criterion of 180 ms to define utterances. Similarly, in Beechey et al. (2018), a 300 ms criterion was used. Using the criterion of 1.25 s as was used by Hadley et al. (2019), we re-analyzed the recordings from the present study and those from Watson et al. (2020) and found that utterance durations decreased in noise (as opposed to increasing when a 180 ms criterion was used). When inspecting the segmentation produced when using a criterion of either 1.25 s or 180 ms, we find that, for our recordings, the shorter criterion is more sensitive at classifying connected vs unconnected utterances. Many conversational floor transfers remain undetected when using the longer criterion, because individual shorter utterances are glued together into a single long utterance. As the conversations in Watson et al. (2020) included both “free conversation” and solving the Diapix task, we speculate that the increase in utterance length in noise observed by Hadley et al. (2019) is due to the much longer criterion used to segment connected utterances.

2.4.3 Overlaps-within

In natural dialogue, utterances do not always alternate between talkers. Sometimes an utterance of one talker occurs temporally completely within an utterance of the other talker, who continues to maintain the floor. We refer to these types of utterances as overlaps-within. The average duration of overlaps-within were around the duration of one syllable (between 228 and 264 ms), and the frequency was quite low (between, on average, 1.68 and 1.95 occurrences/minute per person), which is well in line with the suggestion in Sacks et al., 1974 that interlocutors try to minimize overlaps in conversation.

In L2, the duration of the overlaps-within increased. However, talkers also spoke slower in L2 and the observed increase in the duration of overlaps-within are similar to the increase in change in average syllable lengths between L1 and L2. In L2 the rate at which overlaps-within occurred increased. In L2 as well as in noise, the proportion of overlaps-within that included repeated words or syllables increased. We speculate that this may reflect increased difficulty in speech planning. Further, in noise there were significantly more attempts to take a turn, suggesting that talkers had more difficulty in achieving fluid turn-taking behaviour.

Across all the conversations, over half of the overlaps-within were verbal backchannels, which serve as markers for agreement or other presence feedback. We had expected that when communication became more challenging, verbal backchanneling would increase, as the need for the listener to acknowledge understanding or indicate presence may be increased. However, the proportion of overlaps-within that were verbal backchannels was similar across all four conditions. In the study by Watson et al. (2020) they observed that, when talkers solved the Diapix task they produced shorter IPUs and fewer overlaps-within than when they held free conversation ("small talk"). This suggests that the Diapix task elicits conversations in which the characteristics of overlaps-within differ from those of free conversation. We speculate that when solving the Diapix task, interlocutors can adopt a question/response or statement/affirmation type of interaction reducing the utility of verbal backchannels. This might explain why we did not observe an increase in the proportion of verbal backchannels in the more challenging conditions.

2.4.4 Speech levels and SNR

As anticipated, talkers increased their speech level in the presence of noise. On average, the speech levels estimated 1 m from the talkers (i.e., the levels reproduced over the headphones) were 58.1 dBA SPL in quiet and 67.5 dBA SPL in the presence of the 70 dBA SPL noise. In speech listening studies it is common for experimenters to manipulate the SNR and observe changes in performance or measure the SNR needed to achieve some fixed level of performance such as speech reception thresholds (SRTs). However, the noise level was fixed in the present study. Thus, the SNR of the acoustic signal received by a listener was determined by the speech level of the talker, and the talker was free to adapt this level. In the present study, the conversations in noise were held with an average SNR of -2.5 dB. Other studies involving interactive conversations have observed similar average SNRs during conversations between normal hearing listeners in noise (-2.5 dB in Mansour et al., 2021, -1.64 dB in Weisser and Buchholz, 2019).

Previous listening experiments have found that listeners perform worse on speech intelligibility tasks in L2. For example, Wijngaarden et al. (2002) found Dutch speakers had SRTs that were 1 to 7 dB higher in their second (English) and third (German) languages compared to their native language. Thus, in the present study, we expected higher average SNRs for conversations in L2 compared to L1. However, talkers spoke at the same average levels in L2 as they did in L1, and thus, the SNR was similar between conditions.

A possible explanation for this could be that, in the presence of the 70 dBA SPL noise, talkers were operating near their physical limit, and could not further increase their speech levels to achieve a more favourable SNR in L2. However, since the average level produced in noise was 67.5 dBA SPL, we think this is very unlikely. Other studies have observed that participants are able to produce conversational speech well above 70 dB SPL (e.g., Beechey et al., 2019; Sørensen et al., 2021, submitted; Weisser and Buchholz, 2019). Through an informal listening of the recordings in noise, all of the authors had the subjective impression that participants were capable of further increasing the level of their voice (i.e., they had not reached their physical limit).

All of the participants reported that they were comfortable holding a conversation in English and had participated in at least one course at a university level where the instruction was given in English. The results in Wijngaarden et al. (2002) suggest that there is a relationship between language proficiency and SRT.

Thus, it is possible that the participants in this study were sufficiently proficient in English so as to be equally good at understanding English and Danish speech in noise. However, we observed a decrease in articulation rate of approximately 11% in L2, which suggests they were less fluent in L2 than in L1 (De Jong and Wempe, 2009; García Lecumberri et al., 2017). Further, as in García Lecumberri et al. (2017) and Van Engen et al. (2010), it took the participants longer to solve the task when communicating in L2, and floor transfers occurred at a slower rate. Taken together, this suggests that although the participants may have been highly proficient in English, they were less fluent in English conversation than in Danish.

Previous studies suggest that one of the factors leading to higher SRTs in L2 is that native listeners gain more benefit from linguistic context (Golestani et al., 2009; Mayo et al., 1997). However, in this study, the pictures used in the Diapix task provided the participants with context information independent of the language spoken. The scene (either beach, farm, or street) and the objects present in each picture could aid listeners in a manner similar to that provided by linguistic context. Since these visual cues were equally available when talkers spoke in L1 and L2, their presence may have reduced differences in listening effort across the language conditions.

In the L2 conditions of the present study, both talkers spoke in their second language and shared the same L1. It is possible that the participants in the present study benefitted from a matched accent (e.g., Peng et al., 2016; Van Engen et al., 2010). However, the native-Dutch speakers in Wijngaarden et al. (2002) did not benefit from a matched accent. Their SRTs were slightly higher for matched-accented English than for native-English. In Wijngaarden et al. (2002), native talkers of the participants' L2 "translated" the sentences to have equal complexity across the languages tested. We speculate that in the present study in L2, talkers may have used grammar that was simpler and words that occur with higher frequency than speech produced by native talkers (e.g., Van Engen et al., 2010). This could further reduce differences in listening effort across the language conditions.

2.5 Conclusion

The purpose of the present study was to investigate whether turn-taking behaviour was affected by manipulations to the expected communication difficulty in dialogue. Overall, participants took longer to solve the task both in L2 and in background noise, suggesting these conditions were more difficult. We hypothesised that the increased difficulty of conversing in L2 and/or in background noise would result in more variable and/or delayed timing of turn-taking. In noise, we saw a small increase in the median and IQR of FTOs. In L2, there was a small decrease in the median FTO, but a small increase in the IQR. Overall, while the effects were statistically significant, they were small. In both noise and L2, talkers increased the duration of their IPUs and took fewer turns. Additionally, talkers spoke slower in L2. All of these changes could result in reducing the difficulty for both the listener and talker. Thus, we speculate that talkers either had spare capacity to overcome the difficulty of communicating in L2 and noise, and/or that they adapted to the situation by changing other aspects of their communication behavior.

2.6 Data availability statement

The recordings obtained in this study can be found on Zenodo:
https://doi.org/10.5281/zenodo.1204951.

2.7 Acknowledgment

The results from an initial analysis of the study presented here was reported in Sørensen et al. (2020a). Note that this initial analysis used the same threshold in the VAD for all conversations. As a result, there are some differences between the findings from that initial analysis and the ones here that are based on a more accurate segmentation and categorization of talkers' utterances.

3

The effect of conversational task on turn taking in dialogue[2]

Abstract

In previous studies, several methods have been used to elicit conversation between talkers. Some involved participants solving a shared task (e.g., describing a map or finding differences between two near-identical pictures), while others have recorded more spontaneous dialogue (e.g., telephone calls). Since the goals of the talkers, and thus the definition of successful conversation, varies across these methods, it is thought likely that turn-taking behavior will vary depending on how conversations are elicited. The present study investigated this by eliciting English conversations from 7 pairs of native-Danish talkers using two methods: solving a Diapix task and engaging in unguided "small talk". For each method, in both quiet and 70 dBA babble noise, two conversations were recorded for each pair. Overall, several differences in conversational behavior were observed. When engaged in "small talk", participants spoke more rapidly, produced longer interpausal units (units of connected speech surrounded by silence), and replied more quickly than compared to when they were solving the Diapix task. These within-pair differences indicate that comparisons of behavior across studies should also consider the method by which conversations were

[2] This chapter is based on Watson, S., Sørensen, A. J., & MacDonald, E. (2020). *The effect of conversational task on turn taking in dialogue.* Proceedings of the International Symposium on Auditory and Audiological Research, 7, 61-68. As new analyses have been performed on the data presented in Chapter 2, we are referring to this study rather than Sørensen et al., 2020a, and consequently we edited some of the comparisons in this chapter compared to in Watson et al. (2020). The figures have been adapted to match the figures in the rest of this thesis. For consistency, the term "IPU" rather than "utterance" has been used.

elicited.

3.1 Introduction

Recent studies investigating the effects of noise and hearing loss on interactive communication have suggested conversational effort could be assessed using measures of speech production and turn-taking behavior (Chapter 2, 4 and 5, Beechey et al., 2018; Hadley et al., 2019). However, for some proposed metrics, the pattern of results vary substantially between studies (e.g., interpausal unit (IPU; connected units of speech surrounded by silence) duration increasing in noise for some studies vs. decreasing in others). A possible explanation for this could be differences across studies in the method used to elicit conversations. However, as pointed out in Chapter 2, the time constants used for bridging gaps between neighboring talk-spurts can also influence these differences.

When talkers switch turns (i.e., there is a transfer of who has the floor), the acoustic signals produced by each talker may partially overlap or be separated by a silent gap. The length of this interval (with a negative sign for overlap and positive for gap) is termed the floor-transfer offset (FTO). It has been hypothesized that in conditions where communication difficulty is increased, the FTO distribution should shift to the right when speech planning is delayed due to limited resources (Chapters 2, 4, and 5). In addition, if increased difficulty decreases the saliency of acoustic cues used to predict the timing of turn ends, then the FTO distribution should become more broad.

In the present study, we investigate the potential effect of task on several metrics of speech production and turn-taking behavior when participants were engaged in both free conversation ("small talk") and when solving the Diapix task (Baker and Hazan, 2011), where the participants find differences between two almost identical pictures by describing them to each other.

3.2 Method

Fourteen normal-hearing native-Danish talkers were recruited for the study (mean age 23). They were divided in pairs (3 male-male, 3 male-female, and 1 female-female) and individuals in each pair did not know each other before the experiment. All participants reported normal hearing and were comfortable communicating in English. The procedure was approved by Science-Ethics

Committee for the Capital Region of Denmark (reference H-16036391) and all participants gave informed consent.

During the experiment, participants were seated in separate isolated sound booths and had no visual contact with each other. They spoke into Shure SM35 microphones that were connected with the GLXD15 wireless systems. The microphone signals were mixed using an RME Fireface 802 sound card and presented over Sennheiser HD650 headphones such that each individual heard his/her partner's voice at the same level as if he/she were standing 1m away.

Each pair produced two conversations in each of four conditions: Diapix task in quiet, Diapix task in noise, five minutes of "small talk" in quiet, and five minutes of "small talk" in noise. The noise used in this experiment was a 20-talker babble presented at 70 dBA SPL and was the same as that used in the study presented in Chapter 4. The conversations were recorded in two blocks. In each block, a conversation in each of the four conditions was collected, with the conditions randomized in order.

The recorded conversations were analyzed in the same manner as in Chapters 2, 4, and 5. For each talker, average speech levels, articulation rates, and IPU durations were measured. Here, IPUs are defined as portions of speech that are separated by acoustic silences of more than 180 ms. In addition, two measures related to turn taking were recorded: FTOs and overlaps-within. As described above, the FTO is the interval between when one talker stopped and the other started speaking. However, in natural dialogue, turns do not always alternate between talkers. Sometimes the turn of one talker occurs completely within that of the other talker. We term these overlap-within because the IPU is temporally overlapped within the turn of the other talker, who continues to maintain the floor.

3.3 Results

Articulation rates, averaged across talkers, in each of the four conditions are plotted in the left panel of Figure 3.1. While no effect of noise was observed on articulation rate, talkers spoke more quickly during free conversation. A repeated measures ANOVA confirmed a significant main effect of task [$F(1,107) = 26.445, p < .001$]. No significant main effect of noise [$F(1,107) = 0.171, p = .68$] or significant interaction [$F(1,107) = 0.015, p = .902$] was observed.

Speech levels, averaged across talkers, in each of the four conditions, are

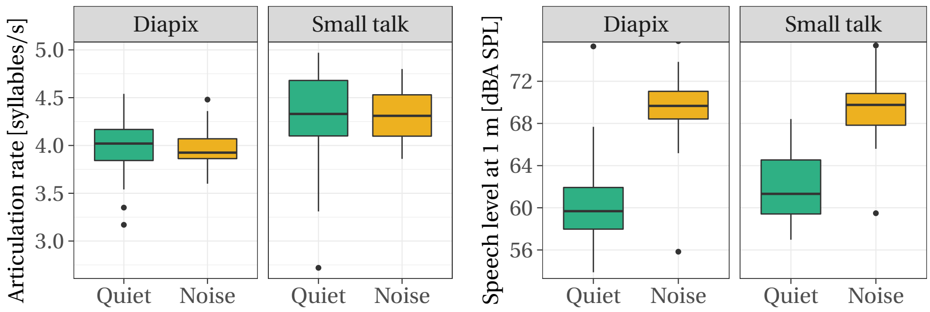

Figure 3.1: Boxplots of articulation rate (left panel) and speech level (right panel) produced by each talker in the four combinations of task and noise. Here and in later plots, the boxplots show the 25th, 50th (median), and 75th percentiles, and the whiskers indicate minimum and maximum observations. Outliers are observations above or below 1.5 times the interquartile range.

plotted in the right panel of Figure 3.1. Consistent with the Lombard effect, talkers increased speech levels in noise. However, speech levels were similar in the two tasks. A repeated measures ANOVA confirmed a significant main effect of noise [$F(1,107) = 117.175, p < .001$]. No significant main effect of task [$F(1,107) = 1.656, p = .201$] or significant interaction [$F(1,107) = 1.1738, p = .19$] was observed.

For every instance where talkers switched turns, the floor-transfer offset (FTO) was calculated. The left panel of Figure 3.2 presents FTO distributions for each of the four combinations of task and noise (presented as kernel density plots of pooled observations across pairs and blocks). The median and interquartile range (IQR) of each talker pair's FTOs are plotted in the upper and lower right panels of Figure 3.2.

From Figure 3.2, it can be seen that task and noise had different effects on the distribution. The median FTO was shorter during small talk than during the Diapix task, but did not change in the presence of noise. A repeated measures ANOVA confirmed a significant main effect of task [$F(1,34.537) = 5.665, p < .001$]. No significant main effect of noise [$F(1,51) = 1.135, p = .442$] or significant interaction [$F(1,51) = 0.014, p = .908$] was observed. In contrast, while the FTO IQR was similar across tasks, it increased in noise. A repeated measures ANOVA confirmed a significant main effect of noise [$F(1,51) = 25.577, p < .001$]. No significant main effect of task [$F(1,51) = 1.008, p = .32$] or significant interaction [$F(1,51) = 0.161, p = .689$] was observed.

The distributions of IPU durations in the four conditions is plotted in the

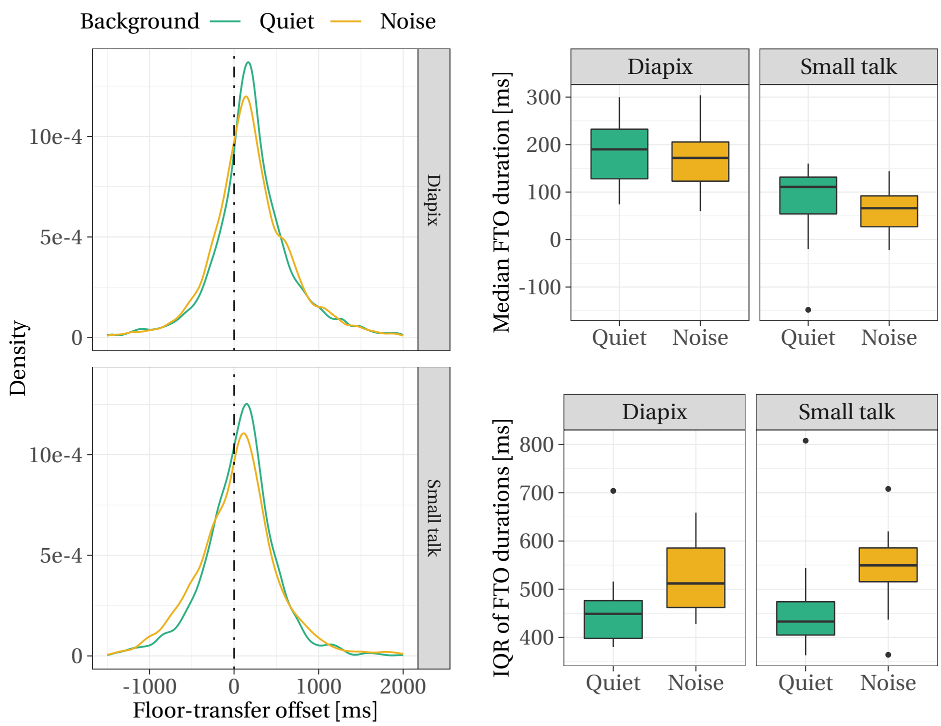

Figure 3.2: Left panel: Distributions of floor transfer offsets (FTO) pooled across pairs and replicates (left panel). Right panel: boxplots of the median (upper right) and interquartile range (lower right) of each pair's FTOs for the four combinations of task and noise.

left panel of Figure 3.3. Note that here, IPUs that were categorized as overlaps-within have been excluded. The median IPU duration increased both in noise and in small talk (see the right panel of Figure 3.3). A repeated measures ANOVA confirmed significant main effects of noise $[F(1,51) = 20.396, p < .001]$ and task $[F(1,51) = 14.79, p < .001]$ and no significant interaction was observed $[F(1,51) = 0.024, p = .877]$.

The rate at which overlaps-within occurred increased in small talk (see Figure 3.4). A repeated measures ANOVA confirmed a significant main effect of task $[F(1,51) = 10.617, p < .01]$. No significant main effect of noise $[F(1,51) = 1.175, p = .28]$ or interaction $[F(1,51) = 0.035, p = .852]$ was observed.

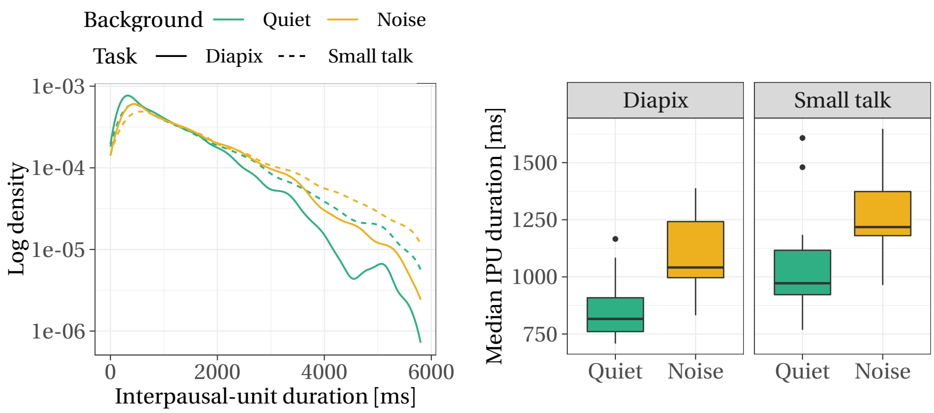

Figure 3.3: Distributions of IPU duration plotted with a logarithmic y-axis (left panel) and boxplots of each pair's median IPU duration (right panel) for the four combinations of task and noise.

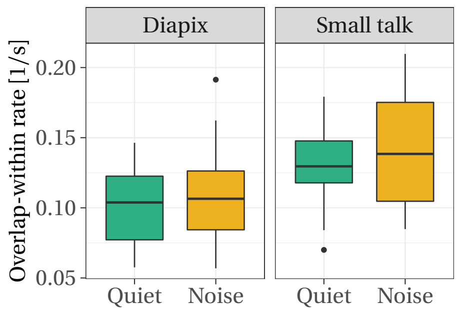

Figure 3.4: Boxplots of the rate of occurrence of overlaps-within (i.e., turns from one talker that occur completely within a turn of the other talker) for the four combinations of task and noise. Note that the rate has been normalized by the total phonation time rather than duration of the conversation.

3.4 Discussion

The purpose of the present study was to investigate if the method of eliciting dialogue between two talkers affected various measures of speech production and turn-taking behavior. Over the course of the study, pairs of talkers, who were not familiar with each other prior to the experiment, produced eight conversations in four different conditions. In half the conditions, talkers were instructed to participate in small talk (i.e., a free conversation). In the other half, they conducted a Diapix task, where they had to find differences between two almost identical pictures. Half of the conversations were conducted in quiet, the other half were conducted in a background of multi-talker babble noise. Overall, changes in speech production and turn-taking behavior were observed across the four conditions. Further, the pattern of results indicated that while both background noise and conversational task influence dialogue behavior, they have different effects.

3.4.1 Speech production

Consistent with the Lombard effect, talkers increased speech levels in the presence of noise, but the levels were not influenced by the task. In contrast, talkers spoke more rapidly when participating in free conversation than when solving the Diapix task. However, their speech rate was not influenced by the noise.

The influence of noise on articulation rate in previous studies of conversation has been inconsistent. While the same Diapix task was used in the studies in Chapter 2 and 4, the normal-hearing talker pairs in Chapter 2 did not change their speaking rates in noise, whereas the normal-hearing talkers in Chapter 4, who conversed with hearing-impaired talkers, decreased their rate of speech when talking in noise, indicating different behavior depending on the conversational partner.

3.4.2 Floor-transfer offset (FTO)

It has been hypothesized that in conditions where communication difficulty is increased, the FTO distribution should shift to the right when speech planning is delayed due to limited resources (e.g. Chapters 2, 4, and 5). Further, if increased difficulty decreases the saliency of acoustic cues used to predict the timing of turn ends, then the FTO distribution should become more broad.

In the present study, the median FTO during the Diapix task was longer than during free conversation. It is tempting to conclude that conducting the Diapix task is more challenging than holding free conversation. However, no change was observed between the quiet and noise conditions. If it was conversational effort that was responsible for the longer median FTO observed when the Diapix task was conducted in quiet, then one would expect that adding noise would further increase the difficulty and result in an even longer median FTO. However, this was not observed.

One possible explanation for these results is that participants are communicating differently between the conversational tasks. To solve the Diapix task quickly may require more accurate information transmission than is needed in free conversation. Thus, talkers might adjust behavior and target a longer FTO to reduce the number of speech overlaps. Another possible explanation is that solving the Diapix task may involve more question-answer constructions than free conversation, some of which may require a visual search to be completed (e.g., “Do you see a red ball?"), delaying the response from a talker.

While the IQR of the FTO distributions increased in noise, there were no differences across conversational tasks. Since the FTO distributions for free conversation and solving the Diapix task were similar in breadth, these results suggest that the ability to predict the timing of turn ends was not influenced by task. The broader FTO distributions observed in the presence of noise are consistent with a reduction in ability to predict the timing of turn ends, which is likely due to a reduction in the saliency of acoustic cues used to make the predictions.

3.4.3 IPU duration

The median IPU duration was observed to be longer during free conversation and also increased in the presence of noise. In Chapter 2 we also observed increased IPU duration in noise and we suggested that this was due to talkers holding their turn longer, providing more time for interlocutors to conduct speech planning and speech understanding.

In that study, the slopes of the distributions of IPU duration were different in quiet vs. noise. However, in the present study, the differences in median IPU duration across conditions appear to be driven mainly by differences in the frequency of very short IPUs (i.e, approximately 500 ms or shorter, which corresponds to 1-2 syllables). For IPU durations ranging between 750-2000ms, the slopes of the distributions are similar across the four conditions. This is consistent with a possibly increase in the number of simple short responses during the Diapix task (e.g., “Yes”, “Uh...”,“Yep”, “Huh...”).

3.4.4 Overlap-within rate

In natural dialogue, turns do not always alternate between talkers. Sometimes the turn of one talker occurs completely within that of the other talker (i.e., it is overlapped within the turn of the other talker who continues to maintain the floor).

In the present study, overlaps-within occurred more frequently during small talk than when conducting the Diapix task. One possible explanation for this is a difference in the conversational goals between small talk and solving a Diapix task. As mentioned above, to solve a Diapix task rapidly, participants should aim to maximize the rate of information transfer. As a consequence, they may attempt to reduce the rate at which they interrupt their partner. In contrast, during small talk, the quality of the social interaction may be prioritized over the rate at which information is transmitted.

However, for free conversation, both longer IPU durations and a shift of the FTO distribution to the left were observed. Thus, it is also possible that the increase in the rate of overlaps-within are a natural consequence of these changes rather than a change in conversational goals.

3.5 Summary

When participating in small talk compared to the Diapix task, talkers spoke more rapidly, produced longer IPUs, produced overlaps-within more frequently, and when a turn switched, the floor-transfer offset was shorter. When holding conversation in noise, talkers increased the level of voice, produced longer IPUs, and the distribution of floor-transfer offsets was more broad.

3.6 Acknowledgment

Parts of this study were supported by the William Demant Foundation (16-3968).

4

Conversational dynamics in task dialogue between normal-hearing and hearing-impaired interlocutors[a]

Abstract

This study investigated the effects of noise and hearing impairment on conversational dynamics between pairs of young normal-hearing (NH) and older hearing-impaired (HI) interlocutors. Twelve pairs of NH and HI individuals completed a spot-the-difference task in quiet and in three levels of multitalker babble. To achieve the rapid response timing typical of conversational turn-taking, people must simultaneously comprehend incoming speech, plan a response, and predict when their partners will end their turn. In difficult conditions, we hypothesized that the timing of turn taking by HI interlocutors would be delayed and more variable. We found that the timing of turn starts by HI participants had a higher variability than NH participants, and both NH and HI participants started turns later and with more variability in the presence of noise. For both groups, there was a negative relationship between the signal-to-noise ratio (SNR) produced by their partner and both the median and interquartile range of timing of when they started their turn. In noise, it took the pairs longer to complete the task, and they spoke louder and slower. They also produced longer interpausal units, i.e., units of connected speech surrounded by silence, which, counterintuitively, can ease speech planning. NH participants adapted

[a] This chapter is based on Sørensen, A. J. M., MacDonald, E. N., & Lunner, T. (submitted) *Conversational dynamics in task dialogue between normal-hearing and hearing-impaired interlocutors.* Manuscript submitted for publication.

their speech rates to match that of their HI interlocutor at higher noise levels and allowed HI participants to speak more, reducing the need for HI participants to listen. We interpreted this outcome as NH participants adapting to reduce the difficulty experienced by their HI interlocutors.

4.1 Introduction

Traditionally, speech understanding and speech production have been investigated in isolation. In speech perception studies, the task often used is to ask participants to repeat back what they heard or provide a response after having listened to a stimulus. In speech production studies, talkers are often asked to read text aloud but for no apparent listener. However, real conversation is not just the sum of production and listening; it is an interaction between two or more participants who use dynamic feedback and adaptation to engage in this verbal dance. Thus, as suggested by Carlile and Keidser (2020), to appropriately test hearing abilities to address the difficulties hearing-impaired (HI) people experience in everyday interactions, one has to measure performance using tasks that are similar to everyday interactions to ensure that similar neural activity is engaged. In the present study, we investigated conversational turn-taking between young normal-hearing (NH) and older HI interlocutors engaging in a Danish-translated version (DiapixDK, Sørensen, 2021) of the DiapixUK task (Baker and Hazan, 2011) in quiet and in multitalker babble noise conditions presented at three different levels: 60, 65, and 70 dBA SPL. We hypothesized that hearing loss and noise interference should increase listening difficulty reducing the resources available for speech planning and reduce the saliency of cues used to predict the end of turns, resulting in delayed and more variable response times. Below, we elaborate on these hypotheses.

4.1.1 Timing of turn-taking in conversational interaction

The fundamental organization of conversational interaction is the switching of turns between interlocutors (conversational partners). Figure 4.1 illustrates the basic conversational states between talkers in dialogue. A switch in the conversational turn is termed a floor transfer, and we measure the floor-transfer offset (FTO) as the duration from when the first person stops talking to when the next person starts talking. We define interpausal units, or IPUs, as sequences of connected speech in which any included acoustic silences are less than 180 ms. In Figure 4.1, there are two floor transfers between IPUs from Talker A and Talker B: the first one happens in an overlap between the talkers (overlap-between), and the next happens in a gap between the talkers. These are pooled to obtain floor-transfer offsets (FTOs), where overlaps-between are negative FTOs and gaps are positive FTOs. Finally, we define pauses as joint periods of silence between talkers that are not followed by a floor transfer, and overlaps-within as joint speech during an IPU of one talker that does not result in a floor transfer.

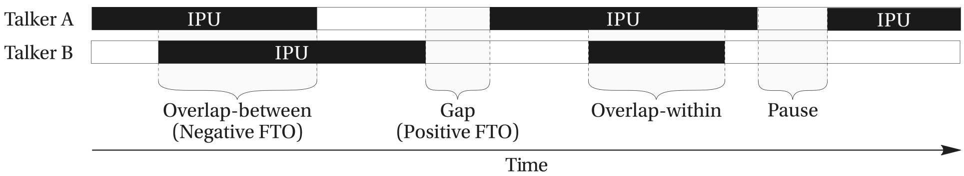

Figure 4.1: Illustration of the classification of gaps, overlaps-within, overlaps-between, pauses, and interpausal units (IPUs) during conversations between two talkers: Talker A and Talker B. There are two floor-transfer offsets (FTOs): the overlap-between and the gap. We define an IPU as a unit of connected speech surrounded by silence of at least 180 ms. A pause is an acoustic silence greater than 180 ms that is surrounded by IPUs from the same person and does not include a floor transfer. An overlap-within is a speech stream occurring completely within the other person's IPU that does not result in a floor transfer. The figure is adapted from Sørensen et al. (in press).

Distributions of FTOs have shown that in dialogue, the typical response time is slightly positive with a peak around 200 ms and that both overlaps and gaps occur during floor transfers (Aubanel et al., 2011; Brady, 1968; Heldner and Edlund, 2010; Levinson and Torreira, 2015; Norwine and Murphy, 1938; Stivers et al., 2009). In their model of conversational turn-taking, Levinson and Torreira (2015) outline the following overlap between comprehension and production involved in initiating a turn while listening to the interlocutor's incoming turn (see Figure 4.3 in Levinson and Torreira (2015), for their model). First, listeners need to receive enough information from their interlocutor's turn to understand the action required from them. As soon as they receive enough information, they start planning their turn (Barthel et al., 2017, 2016; Bögels et al., 2015b; Gisladottir et al., 2015). Next, people start formulating their response while they continue to process the incoming turn. Simultaneously, they monitor their interlocutor's turn for possible syntactic, prosodic, or other cues that signify an upcoming turn end. Single-word preparation has been found to take at least 600 ms, while the preparation of multiword utterances takes over one second (Indefrey and Levelt, 2004; Magyari et al., 2014), so this process must take place at least 600 ms before launching a response. While still monitoring for turn-end cues, immediately before providing a response, people prepare their articulators and typically deliver their response 200 ms after the offset of their interlocutor's turn. Thus, given these latencies involved in taking a turn, the typical response time of approximately 200 ms should not be enough time to include response preparation and articulation, and some overlap between comprehension and production must occur. We expect that if we observe delays and more variability in turn-timing in the presence of noise compared to in quiet and for HI participants compared to NH participants, interlocutors may have 1) had to use more resources when planning their turn and 2) been less sensitive to turn-end prediction cues. Below, we elaborate on these two hypotheses.

4.1.2 Increased cognitive effort when speech is degraded by interfering noise or hearing impairment

Barthel et al. (2016) argue that the preparation of responses in parallel with comprehending the incoming turn is cognitively demanding, as the processing resources used for the production and comprehension of speech can interfere with and use resources from the same neurological system (Menenti et al., 2011; Segaert et al., 2012). There is evidence that to comprehend speech in real-time, one must use predictions about the next word. When a mismatch between the expectation and the actual word uttered is encountered, additional resources are recruited to support the ongoing interpretation process (for a review, see Hagoort and Indefrey, 2014). The Ease of Language Understanding (ELU) model (Rönnberg et al., 2008, 2013) argues that when the sensory input is degraded, as a result of hearing loss or noise interference, there is a high probability that the perceived phonological input will not match stored phonological representations in long-term memory. When that happens, there is a shift from an implicit, automatically controlled process to an explicitly controlled process in which additional resources are recruited to infer meaning from the missing information. Gisladottir et al. (2015) found that sentences with fewer syntax constraints were less predictable and required more resources to be processed than more predictable sentences required. If a person misses some of the words or parts of words being spoken due to hearing loss or background noise, the received content may be less predictable, leading to increased cognitive effort when mismatches are detected. Therefore, cognitive demands may be increased during early response preparation, especially when the speech signal is degraded due to noise interference or hearing loss. As a consequence, we expect to observe delayed and less precise timing of turns if interlocutors exert increased effort in response preparation.

4.1.3 Turn-end prediction cues

The following acoustic cues have been found to be turn-yielding: a drop in loudness; a rising or falling pitch contour; an increase in vocal jitter, shimmer, and noise-to-harmonic ratio (NHR); longer IPUs; and a drawl on the final syllable or the stressed syllable of a terminal clause, the use of stereotyped expressions such as “or something”, and the completion of a grammatical clause (Brusco et al., 2020; Duncan et al., 1972; Gravano and Hirschberg, 2011; Hjalmarsson, 2011). Duncan et al. (1972) proposed a summing-of-cues theory stating that the more turn-yielding cues that are present, the more likely it is that a switch of turns is impending; however, if any turn-holding cues are present simultaneously, a switch of turns is implausible. These turn-holding cues include hand gestures and sustained pitch. The summing-of-cues theory has been supported by evidence from Gravano and Hirschberg (2011) and Hjalmarsson (2011).

Hearing loss has been demonstrated to reduce frequency discrimination, intensity discrimination, and modulation discrimination and to increase speech reception thresholds (Moore, 1996). For NH listeners, noise interference has also been shown to reduce intensity discrimination (Schneider and Parker, 1990) and frequency discrimination (Li and Jeng, 2011). Thus, we hypothesized that for a person with hearing loss and for both NH and HI individuals in noise, the acoustic cues described above, such as changes in frequency (pitch, shimmer, and jitter), intensity (loudness), speech rate (modulation), and the completion of grammatical clauses (because of reduced SRTs), may be less salient. If these cues are harder to utilize to predict the upcoming end of a turn, this could manifest as both a delay and more variability in turn-timing. A delay could be introduced if cues are not perceived or are misjudged, and a person has to react to pure silence. Furthermore, if a person cannot hear everything the other person is saying, the point in time at which they can start preparing their response may be delayed as the critical information to understand the response action may be perceived later and as they have to draw on top-down information to understand the incoming speech. If the cues are less salient, providing their response at the “correct” time would be more difficult, increasing the variability in the timing of the response.

4.1.4 Turn-timing and adaptive behavior

Barthel and Sauppe (2019) found that people planned speech responses in parallel with processing their interlocutor’s turn, despite the increased cognitive demands demonstrated by increased peak and mean pupil dilation responses as well as longer peak latencies. The authors argued that, in conversation, people optimize for fast response times at the expense of increased processing load. Thus, timing a turn seems socially important, and turn-taking has been found to have a universal pattern across languages and cultures (Stivers et al., 2009). Therefore, interlocutors may change other aspects of their speech production and interaction to facilitate smooth turn-taking despite increased processing load. In a setup similar to that of the current experiment, Sørensen et al. (in press) found that pairs of NH interlocutors answered slightly later and with more variability in the presence of background noise. Simultaneously, interlocutors increased the duration of their IPUs, and the authors found indications that for longer IPUs, the FTOs were shorter and more precise. Beechey et al. (2018) and Watson et al. (2020) also found that NH interlocutors lengthened their IPUs in noise. We expect to find that interlocutors will also increase their IPU durations in this study and that HI individuals will increase their durations more than NH participants, likely due to a slowing of speech planning (Brusco et al., 2020) and an increased use of filler words (Clark and Fox Tree, 2002).

Hazan and Baker (2011) found that speakers adapted to the difficulty experienced by their conversational partner. They found that NH interlocutors decreased their speech rates when their interlocutor conversed in the presence of babble noise or heard them through a voice vocoder. When communicating with an NH interlocutor, Sørensen et al. (in press) found that NH interlocutors did not change their speech rates in the presence of noise. We expect, however, to find that NH participants in this study will decrease their speech rates with increasing noise levels to facilitate understanding by their HI interlocutors. Furthermore, in Sørensen et al. (in press), NH interlocutors communicated, on average, at -2.5 dB SNR. As HI individuals are known to have higher speech-reception thresholds than NH individuals (e.g., Nielsen and Dau, 2011), we expect NH interlocutors to speak at higher SNRs in this study to increase speech understanding by their HI interlocutors.

4.2 Methods

4.2.1 Participants

Twelve unacquainted mixed- and same-sex pairs of young normal-hearing (NH) and older hearing-impaired (HI) interlocutors were recruited (9 females, 7 mixed-sex pairs). The NH group (age min, mean, max, std = 23, 26, 30, 2.7 years) had hearing threshold levels below 20 dB HL between 125 Hz and 8 kHz. The HI group (age min, mean, max, std = 65, 73.5, 79, 4.4 years) had symmetrical, mild-to-moderate presbyacusis with N2/N3 audiograms (Bisgaard et al., 2010), and participants in this group were unaided during the experiment. The audiograms for both groups are plotted in Figure 4.2. NH participants had no professional experience talking to HI individuals, and they did not know prior to the experiment that they were going to communicate with a person with hearing loss. When introduced to each other before the experiment, the HI participants did not wear hearing aids; therefore, they did not reveal their hearing status to their NH interlocutors. However, it was evident that there was an age difference between them. All participants provided informed consent to participate in the experiment, and the experiment was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). The participants were compensated for their time.

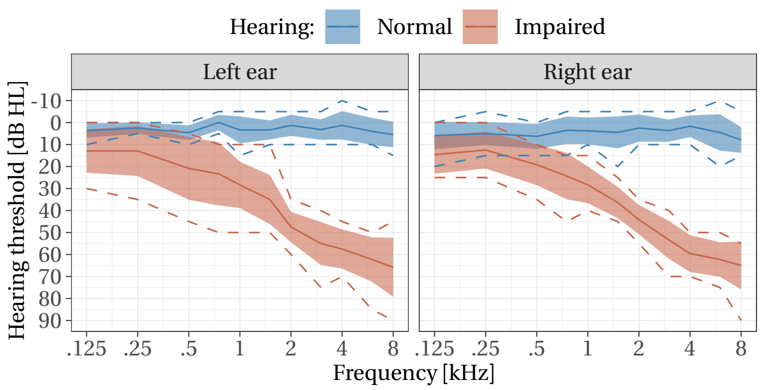

Figure 4.2: Audiometric pure-tone thresholds of normal-hearing and hearing-impaired participants. The solid line indicates the mean hearing threshold, the colored regions indicate one standard deviation, and the dashed lines indicate minimum and maximum measured thresholds.

4.2.2 Setup and presentation level

Participant pairs were placed in separate audiometric booths wearing Shure SM35 head-worn wireless cardioid microphones (transmitted by a Shure GLXD14 wireless system) and Sennheiser HD650 open headphones, with which they communicated with each other. In conditions in which interlocutors were to converse in the presence of background noise, participants heard noise played back over their headphones mixed with their interlocutor's microphone signal. The background noise was 20-talker babble created by taking 20 minutes of recordings from 20 talkers balanced in sex from an NH/NH corpus previously recorded by the author (Sørensen et al., 2018). Pauses were removed using voice activity detection (VAD), and the recordings were normalized to the same RMS level as the recording with the lowest RMS level. Finally, these were added together. The first author listened to the recording and ensured it was impossible to resolve any words from the individual talkers.

The presentation level of the noise in the headphones was calibrated by placing them on a headphone coupler (consisting of a B&K 4149 microphone preamplified by a B&K 2619) connected to a B&K 2636 sound level meter (SLM). Babble noise was played back in the headphones, and the dBA level (10-second integration time) on the 2636 SLM was noted, along with the dB FS digital RMS of the babble snippet played back. The digital gain of the babble was adjusted to reach an SPL of 70 dBA on the SLM.

The levels of the participants' head-worn microphones were adjusted such that the resulting presentation levels over the headphones were the same as if the listener were one meter away from the talker in the same room. This was achieved by adjusting the gain of the head-worn microphone to match that of a reference microphone placed one meter away from the talker.

4.2.3 Task and procedure

To elicit dialogue between pairs, participants were asked to complete the DiapixDK task (Sørensen, 2021), a translated version of the DiapixUK task (Baker and Hazan, 2011). In the experiment, conversational pairs of NH and HI participants solved the pictures from the DiapixDK corpus in each of four conditions: in quiet or in the presence of 20-talker babble background noise (see Section 4.2.2) presented at 60, 65, and 70 dBA SPL. Participants repeated each of the conditions three times, separated into three blocks, and had a break in between each block. The four conditions were randomized within each block. Each Diapix picture appeared an equal number of times in each condition.

Before the test, a training round was conducted outside the booths where the participant pairs sat together, solving a picture pair from the original Diapix task (Van Engen et al., 2010) under the experimenter's supervision. This was performed to familiarize participants with the task. Separated in their booths, the participants had another test round with another picture pair from the original Diapix task in 65 dBA SPL background noise to familiarize them with the setup, the noise, and the procedure. The experimenter sat outside and monitored all conversations. In the test, the participants were given a maximum of 10 minutes to find 10 differences between the Diapix pictures. If they did not complete the task within 10 minutes, the experiment continued to the next condition. In total, 144 conversations (3 replicates × 4 conditions × 12 pairs) were elicited. Of these, 11 were stopped after 10 minutes, but before 10 differences had been found. Overall, the experiment took approximately two to three hours per participant pair, including introduction, training, breaks, and the experiment.

4.2.4 Analysis of recordings

For each of the conversations, voice activity detection (VAD) was performed to obtain a binary speech activity array. The VAD used was energy-based with a window length of 5 ms with 1 ms of overlap. As in Heldner and Edlund (2010), silent portions shorter than 180 ms were bridged to avoid mistaking periods of silence during plosive consonants for actual pauses. Heldner and Edlund (2010) removed bursts of activity shorter than 90 ms as they were assumed to be nonspeech, and Beechey et al. (2018) used a threshold of 75 ms. We found that for our corpus, 70 ms was appropriate to avoid excluding short "yes"-responses but still remove nonspeech elements like short coughs or the impact noise if a person hit the microphone inadvertently. Each threshold was determined for each person in each condition and replicate by hand. This laborious process was performed because it was found that setting a common threshold made the VAD too insensitive. Other more advanced VAD methods were tested, such as detectSpeech from MATLAB's R2020a audio toolbox, based on short-term energy and spectral spread. However, these other methods were often found to miss unvoiced speech portions at the boundaries. As we study timing in turn-taking, having a sensitive boundary estimation is of great importance;

hence, we used the simple energy-based VAD, as it was most accurate for these recordings.

After obtaining binary speech activity arrays, we processed the VAD outcomes further in two steps. First, we used the De Jong and Wempe (2009) Praat script to automatically detect syllables in the conversations using default parameter settings. The script finds syllable nuclei in recordings and computes an articulation rate by dividing the total number of syllables by the phonation time of the recording. As the phonation time is dependent on the VAD, we used our own determined speech boundaries to compute the articulation rates. We extracted TextGrids for all recordings using the Praat script. Next, for each of the talkers in each condition's replicate, a MATLAB script was used to count the number of syllables the Praat script found within the boundaries of our speech activity arrays. In total, there were 34808 utterances (as detected by the VAD), and of these, the Praat script did not detect any syllables in 630 of them. We listened to all of these utterances and found that approximately half were nonspeech elements such as in-breaths, coughs, or noises (291), and the other half included speech with undetected syllables. These were manually counted, and speech portions detected by the VAD procedure that were nonspeech were removed from the binary speech activity arrays.

Next, we listened to all detected portions of speech that had a speech level of less than 50 dBA SPL (377) and kept those that were speech (371), and deleted those that were random noises or nonspeech sounds (6). To estimate the articulation rate, we divided the final number of syllables by the phonation time determined by our cleaned speech activity arrays to obtain an articulation rate.

After cleaning the binary speech activity arrays for each pair, the arrays were fed into a communicative state classification algorithm that categorized the conversations into the states illustrated in Figure 4.1 and explained in the first part of Section 4.1.1.

4.2.5 Statistical procedure

Mixed-effects regression models were fit to the variables in *R* using the *lme4* package. For measures individual for participants, the maximal starting model included fixed effects of background (quiet, 60 dBA SPL noise, 65 dBA SPL noise, and 70 dBA SPL noise), hearing status (normal and impaired), and replicate (1, 2, and 3) with up to third-order interaction. The model also included a random intercept varying among pairs and among persons within pairs. For measures common for the two participants within pairs, we only included fixed effects of background and replicate with interaction as well as a random intercept of pairs. The interaction.plot function from the stats package in *R* was used to determine whether to include any correlated or uncorrelated random slopes of any fixed predictors by pair and by person within a pair. The denominator degrees-of-freedom were Satterthwaite approximated for the F-tests for the fixed effects. The step function in the *lmerTest* package was used to perform backward elimination of factors with a significance level higher than 0.15. The *compare_performance* function from the *performance* package in *R* was used to compare models before and after reduction to pick the model that fit the data best, and residuals plots were used to confirm model assumptions were met. Pairwise comparisons were computed using the *ls_means* function from the *lmerTest* package, comparing least-squares means of the significant effects using the Satterthwaite approximated df.

4.3 Results

A subset of the results from this experiment has been published in Sørensen et al. (2020b). In that conference paper, a common threshold in the VAD procedure was set for analyzing all recordings. In this paper, we individualized the thresholds and processed the VAD results in other ways (see Section 4.2.4). Therefore, the results presented here deviate slightly from those in Sørensen et al. (2020b).

4.3.1 Task completion time

Eleven of the 144 Diapix pairs were not completed within the 10-minute time frame given to finish the task. Of these, eight were in the first replicate, two were in the second replicate, and one was in the third replicate. One was in quiet, three were in 60 dBA SPL noise, three were in 65 dBA SPL noise, and four were in 70 dBA SPL noise. For these 11 cases, we truncated the completion times to 10 minutes.

We added a random intercept of the Diapix picture pair to the starting model, and the final reduced model describing the completion time was defined as follows: completion time ∼ background + replicate + (1 | pair). There was a significant main effect of background [$F(3{,}127) = 4.43$, $p < .01$] and of replicate

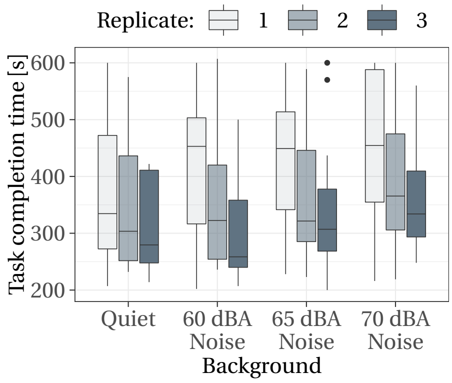

Figure 4.3: Boxplot of the time it took participants to find 10 differences between pictures in the DiapixDK task for each of the three replicates in the four background noise levels: quiet, 60 dBA SPL noise, 65 dBA SPL noise, and 70 dBA SPL noise. Eleven of the 144 Diapix pairs were not completed within the time frame given to complete the task and were truncated to 10 minutes. Here and in later plots, the boxplots show the 25th, 50th (median), and 75th percentiles, and the whiskers indicate minimum and maximum observations. Outliers are observations above or below 1.5 times the interquartile range.

[$F(2,127) = 15.1$, $p < .001$]. A multiple comparison post hoc analysis showed that, on average, there was a significant increase in the completion time of 41 seconds between quiet and 65 dBA [$t(127)$ = -2.31, $p < .05$], of 60 seconds between quiet and 70 dBA [$t(127)$ = -3.43, $p < .001$], and of 41 seconds between 60 dBA and 70 dBA [$t(127)$ = -2.35, $p < .05$]. However, there was no difference between quiet and 60 dBA [$t(127)$ = -1.09, $p = .279$] or between 60 and 65 dBA [$t(127)$ = -1.22, $p = .223$]. There was a significant decrease of 51 seconds between the first and second replicates [$t(127) = 3.32$, $p < .01$], of 83 seconds between the first and third replicates [$t(127) = 5.46$, $p < .001$], and of 32 seconds between the second and third replicates [$t(127) = 2.14$, $p < .05$]. The same analysis was conducted with the 11 unfinished cases excluded, and the pattern of statistical results was the same.

In summary, it took participants longer to complete the task in the two loudest noise background conditions compared to quiet, and there was a learning effect between the replicates.

4.3.2 Speech levels

In the left panel of Figure 4.4, the estimated speaking levels one meter away from the talkers are plotted for each of the four backgrounds for the NH and HI participants. The final model evaluating the speech levels was defined as follows:

speech level ~ background + hearing + background : hearing + (1 | pair/person). There was a significant main effect of background [$F(3,258) = 584$, $p < .001$] and a significant interaction between background and hearing [$F(3,258) = 15.8$, $p < .001$]. A multiple-comparison post hoc analysis showed that there was a significant increase in speech level between quiet and 60 dBA SPL [$t(258) = -20.9$, $p < .001$], between 60 and 65 dBA SPL [$t(258) = -8.21$, $p < .001$], and between 65 and 70 dBA SPL [$t(258) = 11.36$, $p < .001$].

The increase in speech level between backgrounds depended on the hearing status as indicated by the hearing-by-background interaction. There was a borderline difference between NH and HI participants in quiet, where HI participants were estimated to speak 2.3 dBA SPL louder than NH participants [$t(258) = -1.97$, $p = .06$]. While there was no significant difference between the speech levels of the NH and HI participants in the noise conditions, the difference in level depended on the condition. In 60 dBA SPL, HI participants spoke 0.4 dBA SPL louder than NH participants [$t(258) = -.34$, $p = .73$]; in 65 dBA SPL, HI participants only spoke 0.1 dBA SPL louder than NH participants, whereas in 70 dBA SPL NH participants spoke louder than the HI participants by, on average, 0.6 dBA SPL. In the right panel of Figure 4.4, it can be seen that the SNRs produced by both groups depended on the background noise level. The NH and HI participants produced average SNRs of 6.8 dB and 7.2 dB, respectively, in 60 dBA SPL; 3.9 dB and 3.8 dB, respectively, in 65 dBA SPL; and 1.7 dB and 1.1 dB, respectively, in 70 dBA SPL.

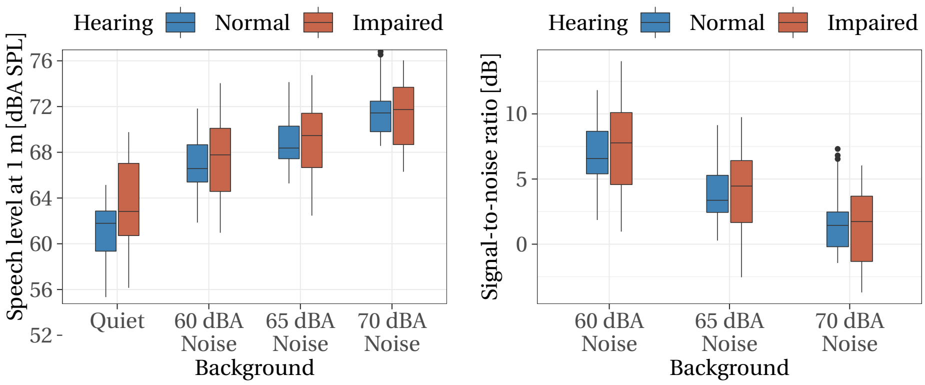

Figure 4.4: Boxplots of speech level (left panel) and signal-to-noise ratios (SNRs, right panel) produced by the normal-hearing and hearing-impaired participants in the four background conditions: quiet, 60 dBA SPL noise, 65 dBA SPL noise, and 70 dBA SPL noise.

In summary, the participants increased their speech levels with increasing background noise level, and the SNR produced decreased with increasing noise level.

4.3.3 Articulation rates

The final model for analyzing the articulation rates was as follows: articulation rate ~ background + hearing + background : hearing + (1 | pair/person). There was a significant main effect of background [$F(3,258) = 23.9$, $p < .001$], a borderline significant main effect of hearing status [$F(1,22) = 4.12$ $p = .055$], and a significant interaction between background and hearing status [$F(3,258) = 2.83$, $p < .05$]. A multiple-comparison post hoc analysis showed that both NH and HI participants had a slower articulation rate in 60 dBA SPL compared to quiet [$t(258) = 5.07$, $p < .001$] and in 70 dBA SPL compared to 65 dBA SPL [$t(258) = 3.12$, $p < .001$], but there was only a borderline difference between 60 and 65 dBA SPL [$t(258) = .158$, $p = .087$]. The extent to which articulation slowed in noise depended on the condition. In quiet and 60 dBA SPL, NH participants spoke 0.3 syllables/second faster than their HI interlocutors ([$t(26.7) = 2.52$, $p < 0.5$] and [$t(26.7) = 2.37$, $p < 0.5$], respectively). However, in 65 and 70 dBA SPL, there was no significant difference in articulation rate between the NH and HI interlocutors ([$t(26.7) = 1.38$, $p = .18$] and [$t(26.7) = 1.46$, $p = .15$], respectively). In summary, the interlocutors spoke slower in noise compared to in quiet, and

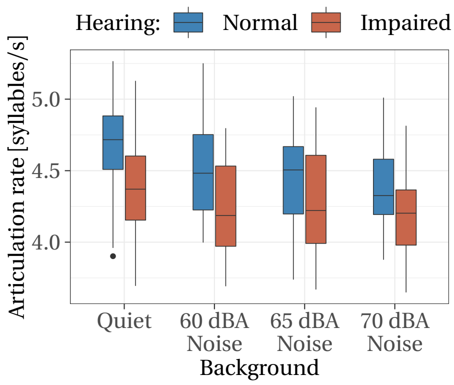

Figure 4.5: Boxplots of articulation rates for normal-hearing and hearing-impaired participants collapsed across all replicates in the four background conditions: quiet, 60 dBA SPL noise, 65 dBA SPL noise, and 70 dBA SPL noise.

interlocutors spoke slower in the loudest noise level, 70 dBA SPL, compared to in 60 and 65 dBA SPL conditions. In the two loudest noise background conditions, there was no difference in articulation rates between NH and HI participants.

4.3.4 Speaking time

As an indication of which interlocutor tended to dominate the conversation, the average proportion of time each person in the two hearing status groups spoke was computed; the interaction effects can be seen in the left panel of Figure 4.6 across the four conditions collapsed across replicates, and across the three replicates collapsed across conditions in the right panel. The proportion of speaking time was measured as the total duration of active speech from the participant (determined by the VAD) divided by the total duration of active speech in the conversation from both participants. As the speaking time of interlocutors always summed to 100%, we performed the statistical analysis on the HI talkers only. To investigate whether the speaking time proportion changed in background noise and with replicates and to determine whether HI participants spoke more than NH participants, we used the following reduced model on the HI subset: speaking time $-50 \sim$ background + replicate + (1 | person). We subtracted 50 from the speaking time in each of the conditions to test whether the values were significantly different from 0, which would indicate that HI participants in those conditions were speaking more than 50% of the time. There was a significant main effect of replicate [$F(2{,}127) = 3.67$, $p < .05$] and a near-significant effect of background [$F(3{,}127) = 2.18$, $p = .09$]. A multiple-comparison post hoc analysis showed that in replicates 2 and 3, HI participants spoke significantly more of the time than in replicate 1 ([$t(127) = -2.36$, $p < .05$], and [$t(127) = -2.34$, $p < .05$], respectively). Furthermore, HI participants spoke 3.6% more in the 70 dBA SPL condition than in quiet [$t(127) = -2.47$, $p < .05$]. Using the *emmeans* package in *R*, we tested whether each condition was significantly different from 0. In quiet, HI participants did not speak significantly more than 50% of the time [$t(12.9) = 1.7$, $p = .11$]. At 60 dBA and 70 dBA SPL, HI participants spoke significantly more than 50% of the time ([$t(12.9) = 2.26$, $p < .05$], and [$t(12.9) = 2.85$, $p < .05$], respectively). For the 65 dBA SPL condition, the difference was borderline significant [$t(12.9) = 2.05$, $p = .06$].

In summary, in noise, HI participants spoke more of the time than NH participants, and the imbalance was more pronounced in the loudest background noise condition of 70 dBA SPL.

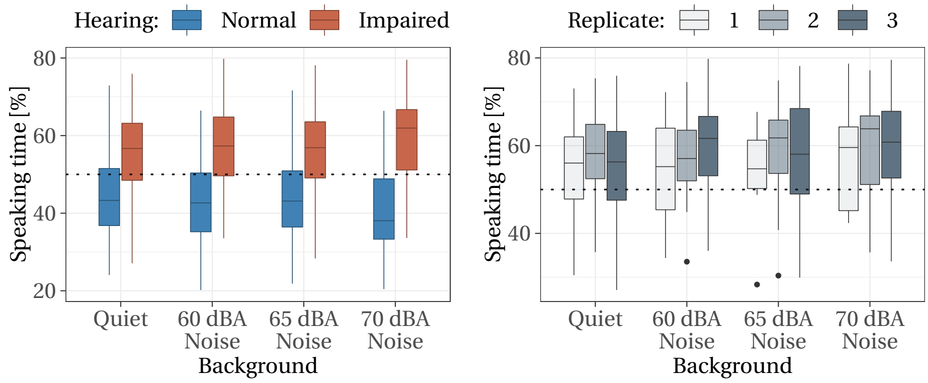

Figure 4.6: Boxplots of the average percentage of time ("speaking time") each group of talkers (normal-hearing, blue, and hearing-impaired, red) spoke in each of the four background conditions: quiet, 60 dBA SPL noise, 65 dBA SPL noise, and 70 dBA SPL noise collapsed across replicates (left panel); for the HI subset, data from the three replicates are combined across background conditions (right panel).

4.3.5 Floor-transfer offsets

Kernel density distributions of FTOs (computed using *geom_density* from the *ggplot2 R* package) collapsed across participants within the NH and HI groups in the four background conditions were computed; see the lower-left panel of Figure 4.7. By visual inspection, the distributions seem broader for the older HI group than for the younger NH group and broader in noise than in quiet.

As a measure of centrality and spread of the FTO distributions, the median and IQR of each person's FTOs in each condition and replicate were computed and plotted in the upper and middle panels of Figure 4.7. The final reduced model was as follows: FTO median $\sim$ background + hearing + replicate + hearing : replicate + $\left(1 + \text{background} \mid \text{pair/person}\right)$. There was a significant effect of background [$F(3,46.2) = 13.5$, $p < .001$] and a significant interaction between hearing and replicate [$F(2,234) = 3.64$, $p < .05$]. There was no main effect of hearing [$F(1,23) = 2.92$, $p = .1$], but a post hoc analysis showed that while there was no difference between NH and HI participants in replicate 1 [$t(26.8) = -.97$, $p = .34$], in replicates 2 and 3, HI participants had borderline significantly longer median FTOs ([$t(26.8) = -1.94$, $p = .062$], and [$t(26.8) = -2.02$, $p = .054$], respectively). A multiple-comparison post hoc analysis showed that there was no difference between quiet and 60 dBA SPL noise [$t(35.5) = -1.75$, $p = .09$], but there was a significant increase of 62 ms between quiet and 65 dBA SPL noise [$t(28.4) = -3.59$, $p < .01$], and a significant increase of 115 ms between quiet and 70 dBA SPL noise [$t(23.1) = -6.7$, $p < .001$]. There was a significant increase between 60 and 65 dBA SPL noise [$t(60.4) = -3.3$, $p < .01$] and between 65 and 70 dBA SPL noise [$t(26.8) = -3.9$, $p < .001$], suggesting that with increasing noise levels, the median increased for both the NH and HI participants.

The IQRs were analyzed with the following reduced model: FTO IQR ∼ background + hearing + replicate + hearing : replicate + (1+background+replicate | pair/person). There were significant main effects of background and hearing ([$F(3,49) = 8.98$, $p < .001$] and [$F(1,22) = 17.9$, $p < .001$], respectively) and a borderline significant interaction between hearing and replicate [$F(2,30.6) = 3.16$, $p = .057$]. Overall, the IQR was larger in noise conditions than in the quiet condition, and with increasing noise level, the IQR also increased. On average, the IQR was 59 ms larger in the 60 dBA condition compared to that in quiet [$t(39) = -3.03$, $p < .01$], 94 ms larger in the 65 dBA condition compared to that in quiet [$t(32.4) = -4.43$, $p < .001$], and 162 ms larger in the 70 dBA condition compared to that in quiet [$t(23) = -4.99$, $p < .001$]. There was a significant increase from 60 to 65 dBA of 35 ms [$t(99) = -2.02$, $p < .05$], and between 65 and 70 dBA of 68 ms [$t(29) = -3.08$, $p < .01$]. On average, HI participants had IQRs that were 152 ms larger than that of NH participants [$t(22) = -4.23$, $p < .001$]. The borderline significant interaction between hearing and replicate was driven by NH participants decreasing their IQRs between replicates 1 and 2 [$t(31.6) = 2.06$, $p < .05$].

The rate of floor transfers was computed jointly for NH and HI participants and can be seen in Figure 4.7, bottom right panel. The final reduced model for analyzing the rate of floor transfers was as follows: floor-transfer rate ∼ background + replicate + (1 | pair). There was a significant main effect of background [$F(3,127) = 22.78$, $p < .001$] and of replicate [$F(2,127) = 8.8$, $p < .001$]. Interlocutors took significantly fewer turns in 60 dBA, 65 dBA and 70 dBA noise levels compared to the number of turns taken in quiet ([$t(127) = 4.57$, $p < .001$], [$t(127) = 5.66$, $p < .001$], and [$t(127) = 8.04$, $p < .001$], respectively), and they took significantly fewer turns in the 70 dBA condition compared to the number of turns taken in the 60 and 65 dBA conditions ([$t(127) = 3.48$, $p < .001$], and [$t(127) = 2.38$, $p < .05$], respectively), but there was no difference between the number of turns taken in the 60 and 65 dBA conditions [$t(127) = 1.09$, $p = .28$]. Participants took significantly more turns in replicates 2 and 3 than in replicate 1 ([$t(127) = -3.26$, $p < .01$] and [$t(127) = 3.91$, $p < .001$], respectively), but there was no difference between replicates 2 and 3 [$t(127) = -0.65$, $p = .52$].

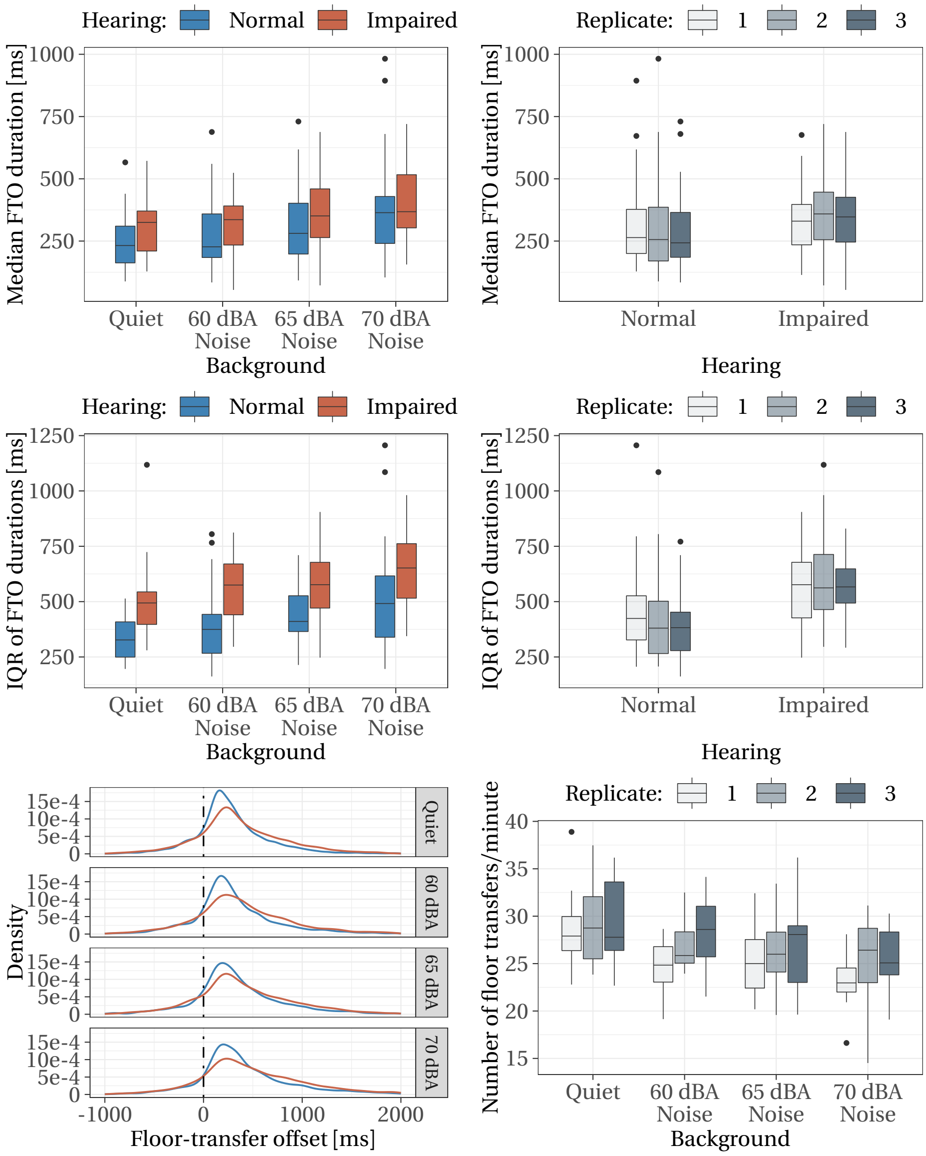

Figure 4.7: Left panels: median floor-transfer offset (FTO) (upper), interquartile range (IQR) of FTO durations (middle), and kernel density plots FTOs (lower) for the young normal-hearing (NH, blue) and older hearing-impaired (HI, red) individuals in the four background conditions: quiet, 60 dBA SPL noise, 65 dBA SPL noise and 70 dBA SPL noise collapsed across the three replicates. Right panels: median of FTOs (upper), IQR of FTOs (middle), and the number of floor transfers per minute (lower) for the three replicates collapsed across the four background conditions. The number of floor transfers in the conversation is summed across participants to estimate the total number of floor transfers in the conversations.

We investigated whether there was a relationship between the SNR produced by a participant's interlocutor and the timing of the participant's turn taking. In Figure 4.8, we plotted the IQR of FTOs as a function of the SNR received by the interlocutor that took over the floor during an FTO. For the median FTOs as a function of the received SNR, see Figure A.3 in Appendix A.1.2. For both the median and IQR of the FTOs of each group, we fitted models of the following form: FTO measure $\sim$ partnerSNRproduced + background + partnerSNRproduced : background + (1 | person). For NH participants, there was a significant negative relationship between their median (-13.6, CI = [-26.7, -0.42]) and IQR (-26.3, CI = [-45.3, -7.4]) of FTOs and the SNR produced by their HI partners ([$F(1,96) = 6.3$, $p < .05$] and [$F(1,38) = 13.13$, $p < .001$], respectively). There was also an interaction between the SNR produced by their partner and the background noise level for the IQR [$F(2,92) = 5.43$, $p < .01$] (and trending for the median [$F(2,92) = 2.55$, $p = .08$]), where the SNR had a larger negative effect on the IQR in 70 dBA SPL noise than in 60 and 65 dBA SPL noise. For HI participants, there was also a significant, negative relationship between the median (-14.2, CI = [-25.9, -2.3]) and IQR (-17.2, CI = [-33.4, -1.05]) of FTOs and the SNR produced by their NH partners ([$F(1,103) = 5.5$, $p < .05$] and [$F(1,67) = 4.24$, $p < .05$], respectively). This suggests that the lower the SNR their interlocutor produced, the harder it was for both NH and HI participants to time their responses.

In summary, the median and IQR of the FTO distributions produced by the NH and HI participants increased with increasing noise levels. HI participants had larger IQRs than NH participants, and there was a trend that their median FTOs were larger than those of NH participants. The participants switched turns at a lower rate in noise conditions compared to quiet and at a lower rate in 70 dBA SPL noise compared to 60 and 65 dBA SPL noise. Compared to the first replicate, in the second replicate, the interlocutors switched turns at a higher rate and the IQR of the FTO distribution of the NH participants decreased. Finally, we found a negative relationship between the median and IQR of the participants' FTOs and the SNR produced by their interlocutors.

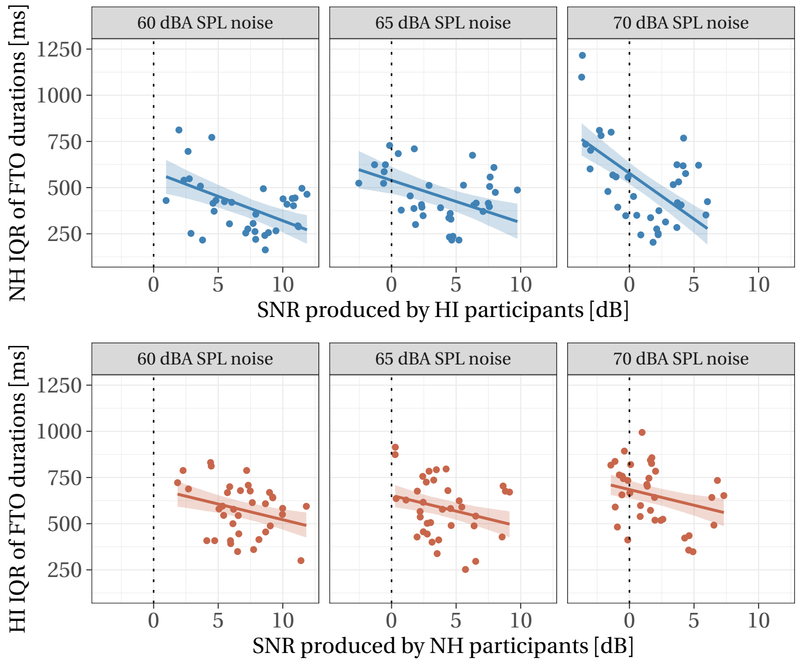

Figure 4.8: Raw interquartile ranges of FTO durations (points) as a function of the SNR received by the interlocutor that took over the floor in the three noise conditions: 60 dBA, 65 dBA, and 70 dBA SPL; data are shown with regression lines and 95% prediction intervals. Data are shown for each participant and replicate for NH participants as a function of the SNR produced by their HI interlocutors (upper panel) and for the HI participants as a function of the SNR produced by their NH interlocutors (lower panel).

4.3.6 Interpausal units

We defined interpausal units, IPUs, as stretches of speech surrounded by 180 ms of silence. These IPUs did not include overlaps-within, which are stretches of speech occurring completely within the interlocutor's turn that do not result in a floor transfer (see Figure 4.1 for a visualization of the conversational categories).

The final, reduced model for analyzing the median IPU durations was as follows: median IPU $\sim$ background + hearing + replicate + hearing : replicate + (1 + replicate | pair/person). There was a significant main effect of background [$F(3{,}235) = 6.99$, $p < .001$] and hearing status [$F(1{,}22) = 8.7$, $p < .01$], as well as a significant interaction between hearing and replicate [$F(2{,}38) = 5.03$, $p < .05$]. A multiple-comparison post hoc analysis showed that there was a significant increase in median IPU duration between quiet and all three noise conditions, where the IPU duration was 95 ms longer in the 60 dBA condition [$t(235) =$ -3.38, $p < .001$], 102 ms longer in the 65 dBA condition [$t(235) =$ -3.63, $p < .001$],

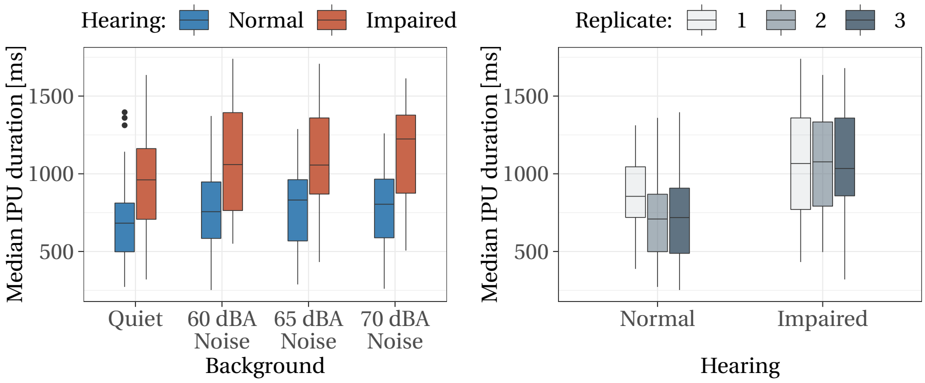

Figure 4.9: Boxplots of median IPU durations for the young normal-hearing and older hearing-impaired individuals in the four background conditions: quiet, 60 dBA SPL noise, 65 dBA SPL noise, and 70 dBA SPL noise collapsed across the three replicates (left panel) and in the three replicates collapsed across the four background conditions (right panel).

and 115 ms longer in the 70 dBA condition [$t(235)$ = -4.08, $p < .001$]. However, there was no significant difference between the IPU duration in any of the noise conditions. On average, HI participants produced IPUs that were 313 ms longer than their NH interlocutors [$t(22)$ = -2.95, $p < .01$]. The interaction between hearing status and background is driven by the decrease between replicates 1 and 2 for NH participants, as visualized in the right panel of Figure 4.9, which was borderline significant [$t(23.8) = 3.55$, p = .071].

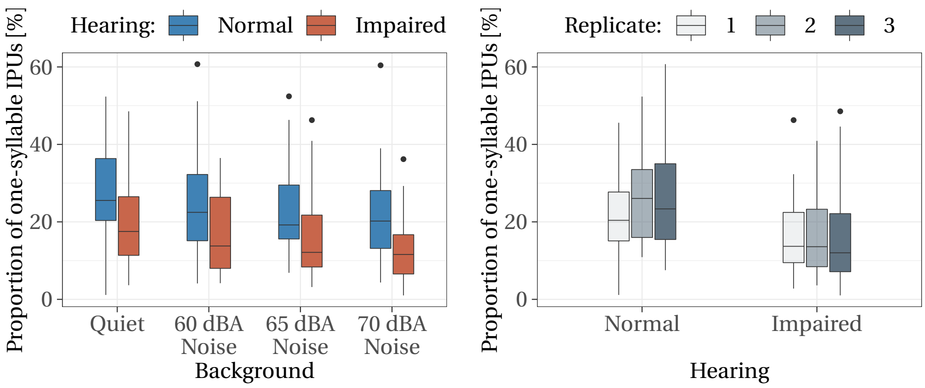

Figure 4.10: The proportion of IPUs that had a duration less than 300 ms, corresponding to one syllable, for the young normal-hearing and older hearing-impaired in the four background conditions: quiet, 60 dBA SPL noise, 65 dBA SPL noise, and 70 dBA SPL noise collapsed across the three replicates (left panel) and in the three replicates collapsed across the four background conditions (right panel).

As a way of investigating whether interlocutors produced more yes/no responses, we investigated the proportion of IPUs that had a duration of one syllable. From the articulation rates, minimum and maximum syllable durations were calculated. The minimum duration was 190 ms, and the maximum duration was 274 ms. We chose a cutoff of 300 ms to ensure that we captured all one-syllable IPUs but were still below the boundary of including two-syllable IPUs. The proportions are plotted for the NH and HI participants in Figure 4.10. The final reduced model for analyzing the proportion of one-syllable IPUs was as follows: proportion ∼ background + hearing + replicate + hearing : replicate + (1 | pair/person). There was a significant effect of background [$F(3{,}257) = 12.8$, $p < .001$] and hearing status [$F(1{,}22) = 5.9$, $p < .05$] and a significant interaction between hearing and replicate [$F(2{,}257) = 4.66$, $p < .05$]. Overall, NH participants produced 9% more one-syllable IPUs than HI participants. In the 60, 65 and 70 dBA conditions, both groups produced fewer one-syllable IPUs than in quiet, and both groups produced significantly fewer in the 70 dBA condition than in the 60 and 65 dBA conditions ([$t(257) = 2.9$, $p < .01$] and [$t(257) = 2.2$, $p < .05$], respectively). NH participants produced significantly more one-syllable IPUs in the second and third replicates than in the first replicate ([$t(257) = -3.23$, $p < .01$] and [$t(257) = -3.84$, $p < .001$], respectively), but there was no difference between replicates for HI participants (see Figure 4.10, right panel).

In summary, HI participants produced longer IPUs and produced fewer one-syllable IPUs than NH participants produced. In the presence of background noise, both NH and HI participants increased their IPU durations and produced fewer one-syllable IPUs. Compared to the IPUs produced in the first replicate, in the second and third replicates, NH participants produced shorter IPUs and more one-syllable IPUs.

4.4 Discussion

We set out to investigate the influence of noise interference and hearing impairment on conversational dynamics between young NH participants and older HI participants when they were completing the DiapixDK task, a spot-the-difference task. In summary, compared to the time taken in the quiet condition, the time it took interlocutors to complete a Diapix picture pair increased in the two loudest noise background conditions. In noise, both NH and HI participants spoke louder and slower, produced longer IPUs, switched turns at a reduced rate, and the timing of the start of their turns was more delayed and had a larger variability. The SNR that interlocutors produced and their articulation rates decreased with increasing noise levels. There was a negative relationship between the median and IQR of a participant's FTOs and the SNR produced by their interlocutor. Compared to NH participants, HI participants produced longer IPUs, and the timing of the start of their turns was more delayed in replicates 2 and 3 and had a larger variability. Under noise conditions, HI participants spoke more of the time than NH participants, and the imbalance was more prominent in the loudest background noise condition. NH participants produced more one-syllable IPUs than HI participants, and both groups produced more one-syllable IPUs in quiet than in noise conditions. NH participants adapted their articulation rates to that of their HI partner in the two loudest background noise levels. Furthermore, NH participants changed their strategy between the first and second replicates; in the second replicate, NH participants produced shorter IPUs with a higher proportion of one-syllable words and spoke less of the time. As a result, their response times became less variable, and the rate of floor transfers increased between the first and second replicates. Below, we elaborate on the implications of these observations.

4.4.1 Timing of turn-taking

Our primary hypothesis was that when NH and HI participants experienced difficulties communicating due to noise interference and hearing loss, they would produce larger and more variable FTOs. Indeed, we found that both the median and IQR of FTO durations increased in noise conditions compared to the quiet condition and both of these measures increased when the background noise level increased. Furthermore, the variability was elevated for HI participants compared to that of NH participants; also, in the second and third replicates, the medians of the FTO distributions produced by HI participants were larger than those of NH participants. We interpret these outcomes as indications that HI participants experienced increased difficulty compared to NH participants and that both NH and HI participants experienced an increase in difficulty in noise. The increase in difficulty may have led to an increase in speech processing load and/or a decreased sensitivity to turn-end cues, altering the distribution of FTOs.

We found a negative relationship between the median and the IQR of FTO durations and the SNR produced by the partner of the participant, indicating that the lower the SNR was, the harder it was for participants to time the start of their turns. There is a higher chance that participants will miss more of their interlocutor's speech at more negative SNRs because of reduced audibility. As discussed in the introduction, inferring meaning from missing information requires more cognitive resources, and we hypothesized that this process of inferring information would lead to delayed and more variable timing of turn taking. Moreover, we hypothesized that the cues used for predicting turn-ends could be reduced for HI participants and in the presence of noise. We interpret the negative relationship between the median and IQR of FTOs and the SNR produced by the participants' partners as indications supporting these hypotheses. Interestingly, in 65 dBA SPL noise, none of the NH participants spoke at negative SNRs, whereas there were a few HI participants who spoke at negative SNRs (see Figure 4.8). Especially in 70 dBA SPL, there were more observations of negative SNRs produced by HI participants than by NH participants, and this outcome had a larger impact on the timing of the NH participants' turns, suggesting that when the noise was louder than their interlocutor's speech, it became more difficult for participants to time their responses. It is well known that understanding speech in negative SNRs is more difficult than in positive SNRs. The reason for the observations of larger FTO variability for NH participants may be due to their interlocutor producing lower SNRs, reducing the audibility of the cues the NH could use for turn-end predictions, and potentially increasing the number of missed words, leading to a higher cognitive load for NH participants who had to infer meaning from the context. NH participants likely produced speech at more positive SNRs to compensate for their partner's impaired hearing, as will be elaborated in Section 4.4.5.

In the upper and middle panels of Figure 4.7, it is visually apparent from the outliers that one of the NH participants had much larger FTOs than the NH average. It appears that this particular NH participant struggled to understand her HI partner, especially in 70 dBA SPL noise. On average, the HI participants in this study produced SNRs of 7.2 dB, 3.8 dB, and 1.1 dB at 60, 65, and 70 dBA SPL noise levels, respectively. However, the HI participant in this pair only produced SNRs of 2.3, -0.7, and -3.2 dB in 60, 65, and 70 dBA SPL noise levels, respectively, i.e., well under the average SNR. The HI participant even spoke faster in 65 and 70 dBA SPL noise levels than in quiet and at a noise level of 60 dBA SPL (4.68 and 4.61 syllables/second vs. 4.50 syllables/second, respectively). Previous studies have found that extended FTOs are associated with dispreferred responses (Bögels et al., 2015a; Kendrick and Torreira, 2015). Informally, when the first author listened to the conversations of this particular pair, she perceived them as “off” and out-of-sync. We speculate that extended gaps not only signal dispreferred responses but also signal difficulty in conversation, as is also suggested by Mertens and Ruiter (2021).

4.4.2 Articulation rates

Both NH and HI interlocutors spoke slower in noise compared to quiet, and in the loudest noise background, they spoke slower than in the quieter noise backgrounds. We found that HI participants spoke slower than NH participants in quiet and 60 dBA SPL, but there was no difference between NH and HI interlocutors in the two loudest noise background conditions. The slowing of articulation rates could be due to a slowing of speech planning, allowing themselves more time to prepare their responses. The slower speech of HI participants may also be a strategy to signal to their NH interlocutors to slow down articulation to facilitate speech understanding. The decrease in articulation rates for NH participants could be an adaptation to the difficulty experienced by HI partners. Piquado et al. (2012) investigated the effect of hearing impairment on sentence recall in younger HI listeners with mild-to-moderate hearing loss compared to age-matched NH listeners. They found that NH listeners had significantly better recall than HI listeners when listening to speech at a normal pace, even though the sentences were adjusted for audibility. However, when given the ability to pause the recording at every clause or sentence boundary, HI listeners performed as well as NH listeners. This suggests that pausing can help HI individuals alleviate their decreased working memory performance, likely because they have a longer time to draw on top-down information to fill in the gaps of what they could not hear. Thus, it is possible that the slowing of the articulation rate by NH participants in the two loudest noise background conditions helped alleviate the increased cognitive load of HI participants when processing their utterances. Hazan et al. (2018b) found that older adults spoke slower than younger adults regardless of the older adults’ hearing status. Thus, the difference we see between the NH and HI groups in quiet and 60 dBA SPL background noise could be driven by age effects rather than hearing status. However, in the two loudest noise background conditions, NH participants slowed down to talk at the same rate as their HI interlocutors. When talking to an NH partner in Sørensen et al. (in press), interlocutors did not slow down in 70 dBA SPL background noise. Beechey et al. (2018) even found that NH interlocutors spoke faster in noise when communicating with an NH partner. Since the NH interlocutors slowed down in noise in the present study, this adaption is likely related to their partners' ability to hear and understand in noise (which can be influenced by both age and hearing loss). This is supported by Hazan and Baker (2011) who found that speakers decreased their speech rates when their interlocutor was receiving a degraded version of their speech. While it could be possible that an age effect could drive the offset in articulation rate between NH and HI participants in quiet and the 60 dBA SPL condition, we speculate that the difference is driven by HI participants exhibiting more communication effort than NH participants, given that HI participants show indications of increased effort compared to NH participants even in quiet, indicated by the larger IQRs of the FTO distributions and longer IPUs.

4.4.3 Interpausal units

We found that in noise, both NH and HI interlocutors lengthened their IPUs (i.e., units of connected speech surrounded by silence), and that HI interlocutors produced longer IPUs than their NH interlocutors. Beechey et al. (2018), Sørensen et al. (in press), and Watson et al. (2020) also found that NH interlocutors lengthened their IPUs in noise in conversations with other NH interlocutors. Brusco et al. (2020) found that the duration of IPUs was longer for IPUs preceding turn-holds than for IPUs preceding turn-switches and backchannels and that people tended to speak slower before turn-holds. The authors interpreted this as planning slowing down speech before holds. If people have not had proper time to plan their turn before launching a response within a socially appropriate time, planning may continue while they are producing their response. In return, individuals may need a longer time to process what they are going to say when they have a hearing loss or are communicating in the presence of background noise. Clark and Fox Tree (2002) found that upon detecting a delay in one's planning of a sentence, people use filler words such as "uh" for the prediction of short delays and "um" for long delays to signal they want to take the turn or continue an ongoing turn. Slowing speech and increased use of filler words would naturally increase IPU durations. While not examined in this study, we speculate that people increase their use of filler words in more challenging conditions. Inserting filler words at the phrase boundary may be a strategy to maintain a socially appropriate timing of turns if more time for speech planning is needed. Beechey et al. (2018) argued that NH interlocutors adopted a "holding the floor" strategy in noise by increasing the duration of IPUs and overlapping more, as it reduced the need to listen to and comprehend their interlocutor's speech. We found that HI interlocutors produced longer IPUs than NH interlocutors, which is well in line with this hypothesis, as it would reduce the listening demands for the person who is experiencing the greatest difficulty performing this task. As further support for this hypothesis, we found that HI interlocutors tended to speak more than the NH did, especially at the highest noise level. Dominating a conversation is a well-known strategy for HI individuals to adopt (Jaworski and Stephens, 1998; Stephens and Zhao, 1996).

4.4.4 Normal-hearing participants changed their strategy after the first replicate

Between the first and second replicates, NH participants changed their strategy. They produced shorter utterances, more one-syllable words, and spoke less of the time, and in return, the interlocutors took more turns per minute. This can be taken as a suggestion that participants adopted a leader/follower communication style, where the HI participant was the leader and the NH participant was the follower, responding with yes/no answers and more rarely producing longer utterances. As the task that they were to complete forced both participants to contribute to the conversation, these effects would likely have been even more pronounced in free conversation. This change in behavior may ease communication for HI participants, as it reduces their need for listening. However, it also increases the responsibility for HI participants to find the objects they need to discuss. Their task-completion time decreased from the first to the second replicate in this study, which was also found in Sørensen et al. (in press). There was an apparent training effect where interlocutors more quickly identified the types of objects that would be different between the pictures (e.g., signs, colors). However, the decrease in completion time could also indicate increased communication efficiency resulting from the more asymmetric contribution between the interlocutors. On average, NH participants also became better at timing their responses in the second and third replicates than in the first, exhibited by the decrease in the IQR of FTOs; this outcome likely occurred because they had to prepare more one-syllable words, which require fewer resources compared to planning multiword utterances.

4.4.5 Speech levels

Under noise conditions, NH and HI participants produced speech levels that were not significantly different from each other. However, the average level difference between the NH and HI participants differed between conditions. In 60 dBA and 65 dBA conditions, NH participants spoke 0.4 and 0.1 dB softer on average, respectively, whereas at 70 dBA, NH participants spoke 0.6 dB louder than their HI interlocutor. This may indicate that the extent to which interlocutors raise their voice levels may depend not only on the background noise level but also on the hearing status of the person. Indeed, Beechey et al. (2020b) found a positive relationship between HI participants' high-frequency average hearing loss and the speech level of their NH interlocutor. This suggests some adaptive behavior from NH participants to the difficulty experienced by HI interlocutors, and conversely, HI participants may be signaling their difficulty by increasing their speech levels. HI participants likely also increased their speech levels due to their reduced audibility of their own voice.

In a similar setup to this experiment, in Sørensen et al. (in press) pairs of NH friends solved the Diapix task together. On average, they communicated at -2.5 dB SNR in 70 dBA SPL background noise. The type of noise used in that experiment was 6-talker speech-shaped noise (ICRA7; Dreschler et al., 2001). In Watson et al. (2020), unacquainted native-Danish NH participants solved the Diapix task in their second language, English, in a setup similar to the present study and with the same type of noise. In that experiment, participants communicated at -0.3 dB SNR, i.e., 2.2 dB higher than those in Sørensen et al. (in press). Compared to Sørensen et al. (in press), the babble used in Watson et al. (2020) had more energy in the 1-3 kHz band, which is detrimental to speech perception. Therefore, some of the SNR differences likely come from a need to produce a higher vocal level to reduce the energetic masking effect. However, the participants in Watson et al. (2020) did not know each other before the experiment and it is possible that this contributed to the increased SNR that was observed. In this study, when solving the Diapix task with an HI conversational partner, NH participants on average produced an SNR of 1.7 dB. This is 4.2 dB higher than that observed in Sørensen et al. (in press) and 2 dB higher than that reported in Watson et al. (2020), suggesting a change in strategy from communicating at negative SNRs when interacting with an NH individual to communicating at positive SNRs when interacting with an HI individual. This outcome was expected, as HI listeners have been shown to have higher SRTs than NH listeners (Nielsen and Dau, 2011). The results of the present study demonstrate that it is possible for NH individuals to communicate at more favorable SNRs than was observed in Sørensen et al. (in press) and Watson et al. (2020). However, in those studies, NH individuals likely did not need to increase their vocal effort beyond what was observed to communicate effectively, as all participants had normal hearing. As argued by Hazan and Baker (2011), people aim to reduce their effort to the minimum that is required to obtain effective communication. This is also supported by our observation that with increasing noise levels, the SNR decreased for both the NH and HI talkers, in line with Weisser and Buchholz (2019) and Pearsons et al. (1977). This supports that people exhibit a trade-off between physical effort and comprehension.

4.4.6 Age vs. hearing loss

The HI participants recruited for this study were much older than the NH participants in this study. Thus, age, in addition to hearing loss, may also have had an impact on our results. In addition to potential age-related changes to hearing, the difference in age between the two groups may have created an environment with different social norms than would be present for conversations between talkers of a similar age. The aim of the present study was to investigate the effect of noise on the timing of turn taking in HI listeners. This study was not designed to disentangle the effects of hearing loss and age. We recruited participants to be homogeneous within hearing status groups to reduce variability from different degrees of hearing loss. However, a small age span within the HI group allowed us to perform some preliminary investigations of the relationship between our objective measures and the age of the participants. We added age as a predictor in all the mixed-effects models but found no statistically significant effect of age on any of the objective parameters we analyzed in this paper. However, this null result should not be taken as evidence against age as a factor. Further work is needed to disentangle any potential effects of age and hearing loss.

4.5 Conclusion

In this study, we investigated the effects of noise and hearing impairment on conversational dynamics between pairs of NH and HI interlocutors. The interlocutors were separated from each other in two booths while they completed a spot-the-difference task, either in quiet or in the presence of multitalker babble presented at three noise levels: 60, 65, and 70 dBA SPL. Talkers decreased their speech rates and increased their speech levels with increasing noise levels. We found that in noise, the medians and IQRs of the FTO distributions of both NH and HI interlocutors increased, suggesting an increased cognitive load and/or a decreased sensitivity to turn-end prediction cues. These effects were even more pronounced for HI individuals than for NH individuals. Furthermore, the pairs increased their IPU durations, and the HI individuals increased their IPU durations more than the NH individuals did, suggesting a delay or slowing of speech planning. We observed indications of adaptive behavior from NH individuals, who reduced their speaking rates and allowed the HI individuals to speak more in the loudest noise background condition. We interpreted this result as an indication that NH individuals aimed to alleviate the difficulty that the HI individuals were experiencing while communicating in noise. Studying interactions between people is of importance when trying to understand the difficulties HI people experience in everyday communication and to help develop hearing assistive technologies that can alleviate some of their difficulties. The results of the current study suggest that measuring some speech measures, such as the timing of turn taking, could be a promising approach for unobtrusively measuring communication difficulty when talkers are exposed to challenging conditions.

5

The effects of hearing aid amplification and noise on conversational dynamics between normal-hearing and hearing-impaired talkers[a]

Abstract

There is a long-standing tradition to test hearing-aid benefits in lab-based speech intelligibility tests, recently accompanied by various cognitive performance measures. In a more everyday-like scenario, the current study investigated the effects of hearing-aid amplification and noise on face-to-face communication between two interlocutors. A total of 11 pairs, consisting of a younger normal-hearing (NH) and an older hearing-impaired (HI) participant, solved spot-the-difference tasks while their conversations were recorded. In a two-block randomized fashion, the pairs solved the task either in quiet or noise, with or without the HI participant receiving hearing-aid amplification. With the addition of 70 dB SPL babble noise, it took participants longer to solve the task, and they spoke louder, had longer interpausal units (IPUs, stretches of speech surrounded by silence), reduced their articulation rates, and had fewer, slower, and less well-timed turn-takings, all indicative of increased conversational effort. The HI produced longer IPUs than their NH interlocutor, and in the presence of background noise, the HI participants had more variable turn-taking timing, and their NH interlocutor spoke louder. When the HI participants were provided with hearing-aid amplification, their turn-timing was less variable, they spoke

[a] This chapter is based on Petersen, E. B., MacDonald, E. N., & Sørensen, A. J. M. (submitted). *The effects of hearing aid amplification and noise on conversational dynamics between normal-hearing and hearing-impaired talkers*, submitted for publication.

faster, and they produced shorter IPUs. In conclusion, measures of conversational dynamics showed that background noise increased the communication difficulty, especially for the HI participants, and that providing hearing-aid amplification caused the HI participant to behave more like their NH conversational partner, especially in quiet situations.

5.1 Introduction

Hearing assistive devices, such as hearing aids, aim to compensate for hearing loss by presenting an amplified and processed acoustic signal to impaired ears. Traditionally, the development of new processing features and designs has focused on the reception of speech and has been evaluated using listening tests of speech intelligibility. These tests either measure the percentage of words or sentences correctly identified at fixed levels of signal-to-noise ratios (SNRs) or adaptively vary the SNR to obtain a certain performance level (e.g., measuring a speech reception threshold). Common for these tests is that they solely focus on the auditory aspect of listening, while at the same time usually being deployed at intelligibility levels where speech understanding can be easily improved, traditionally at a speech reception threshold of 50%. Recently, both the academic and industrial communities have increased the focus on developing and using more ecologically valid testing methods (for a review, see Keidser et al., 2020).

In everyday life, speech intelligibility is not easily measurable but can be assumed to be near-perfect in most communication situations. Attempts have been made to mimic more realistic listening scenarios by improving the SNRs of the speech intelligibility tests to near-perfect performance levels while shifting the focus away from performance as the primary experimental outcome. Instead, a number of alternative measures, usually focusing on the experienced listening effort, becomes the main focus. Item recall (Kuk et al., 2021; Lunner et al., 2016; Sarampalis et al., 2009), changes in pupil dilation (Wendt et al., 2017), and alterations in the brain responses (Mirkovic et al., 2019; Strauss et al., 2010) are some of the alternative measures used to quantify the benefits of hearing-aid processing in more favorable listening conditions. However, despite being more ecologically valid concerning speech intelligibility, these studies have not focused on the fact that listening is seldomly done in isolation. In real-life communication, people engage in conversation involving overlap between speech comprehension and production. In conversation, there is a shared context between interlocutors (conversational partners), and people have the opportunity to conduct repairs, ask for clarifications, and signal they are having difficulty that may lead interlocutors to adapt their way of communicating. While hearing-impaired (HI) listeners often complain about not being able to hear, their difficulties often manifest in communicative interactions because poorer hearing results in miscommunication (Kiessling et al., 2003). Miscommunications cause changes in the conversational dynamics for both the HI listener and their interlocutor. This phenomenon has been investigated in a few studies, showing that when communicating with a HI interlocutor, their normal-hearing (NH) interlocutor adapted the level and spectral content of their speech in a face-to-face conversation (Beechey et al., 2020b; Hazan et al., 2019). Another study found that when seated in two different rooms, i.e., without being able to see each other, the NH interlocutors increased their speaking levels leading to positive SNRs (Sørensen et al., 2021, submitted) compared to when NH were communicating with NH interlocutors, where they communicated at negative SNRs (Sørensen et al., in press). They also found that with increasing noise levels, the NH adapted their articulation rates to their HI interlocutors, and they let the HI dominate the conversation more. The adaptation of the NH talker in these studies was interpreted as signs of trying to alleviate the HI interlocutor from some of the communicational load posed on them by their hearing impairment and background noise.

In order to sustain fluid conversation, interlocutors have to simultaneously comprehend what the other person is saying while planning their own response (for a review, see Levinson and Torreira, 2015. It has been argued that the point of turn-taking between people in a conversation is cognitively demanding because those processes partly take up some of the same cognitive resources (Barthel and Sauppe, 2019; Hagoort and Indefrey, 2014). Whether face-to-face or seated separately, it has been found in a variety of different studies that the typical floor-transfer offset (FTO), i.e. the interval from when one person stops talking to the next person responds, is usually slightly positive with modal response times around 200 ms in dialogue (Aubanel et al., 2011; Brady, 1968; Heldner and Edlund, 2010; Levinson and Torreira, 2015; Norwine and Murphy, 1938; Stivers et al., 2009). This timing of turns has been found to have a universal pattern across languages and cultures (Stivers et al., 2009), and it has been suggested that people chose to optimize for socially appropriate timing of responses at the expense of increased cognitive effort (Barthel and Sauppe, 2019). Because of the time courses involved in word preparation (at least 600 ms for one word, and over one second for multiple words (e.g., Indefrey and Levelt, 2004; Magyari et al., 2014), to meet this rapid response timing people must predict the content of their interlocutor's speech to start planning their response in overlap with the ongoing turn (Barthel et al., 2016; Bögels et al., 2015b; Corps et al., 2018b; Gisladottir et al., 2015; Levinson and Torreira, 2015), and must predict the end of their interlocutor's turn to launch their prepared response at the right time utilizing different turn-end cues (Bögels and Torreira, 2015; Brusco et al., 2020; De Ruiter et al., 2006; Gravano and Hirschberg, 2011). When the acoustic input in a conversation is degraded, such as due to background noise or hearing loss, people can experience a mismatch between their expectations to the unfolding turn and the sensory input. This could require additional resources to infer meaning from the missing information (for reviews, see *The Ease of Language Understanding* (ELU) model in Rönnberg et al., 2013, or Hagoort and Indefrey, 2014). In return, this could result in a delay in the planning of responses. The extent to which the signal is degraded can be variable depending on the predictability of the content, which could result in more variable turn-timing. Moreover, the saliency of the cues used for predicting turn-ends may be reduced for a degraded acoustic signal, resulting in people launching their responses at later and more variable times. Sørensen et al. (2021, submitted) found that with increasing noise levels, both NH and HI interlocutors started their turns later and with more variability, and the turn starts of HI were even later and more variable compared to NH talkers. Sørensen et al. (in press), too, found that NH interlocutors started turns slightly later and with more variability in background noise compared to quiet. Together, these suggested that when exposed to challenging conditions, the changes in FTOs can be used as a measure of difficulty in conversation (communication effort). Yet, compensatory behavior has been observed where people increased the duration of their utterances in noise (Beechey et al., 2018; Sørensen et al., in press, 2021, submitted; Watson et al., 2020),and it has been speculated that, through the use of filler words, interlocutors may still attempt to start their turns at a socially appropriate time even if more speech planning is needed (Barthel and Sauppe, 2019; Sørensen et al., in press, 2021, submitted), and this leads to longer interpausal units (IPU, i.e., stretches of connected speech surrounded by pauses). Thus, the effects of communication difficulty on the median FTO distributions may not be as large as the effects on the variability of FTOs.

The goal of the present study was to investigate whether measures of conversational dynamics can be used to evaluate the effects of hearing-aid amplification and background noise in face-to-face communication between a younger NH and an older HI interlocutor. Participant pairs solved the DiapixDK task (Sørensen, 2021; a Danish-translated version of the DiapixUK task: Baker and Hazan, 2011), , a collaborative spot-the-difference task. As adding background noise to the communication situation affects both talkers, we hypothesized that this will have large effects on the communication. It is expected that the addition of background noise will make the communication especially trying for the HI talkers. We hypothesized that more moderate effects of hearing-aid amplification will be observed, as this alteration is only directly affecting the HI talker. In the present study, we expected that the increase in communication effort induced by adding background noise or not providing hearing-aid amplification will cause the following alterations in the conversational dynamics. **(1)** Longer task completion times (Hazan and Baker, 2011; Sørensen et al., in press; Sørensen et al., 2021, submitted). **(2)** HI participants taking up more of the speaking time (Jaworski and Stephens, 1998; Sørensen et al., 2021, submitted; Stephens and Zhao, 1996). **(3)** Slower and more variable turn-taking (Aubanel et al., 2011; Sørensen et al., in press;Sørensen et al., 2021, submitted). **(4)** Slower articulation rates (Hazan et al., 2018b; Tuomainen et al., 2019) and louder speech (Beechey et al., 2018, 2020a; Sørensen et al., in press; Sørensen et al., 2021, submitted; Watson et al., 2020). **(5)** Longer IPUs (Beechey et al., 2018; Sørensen et al., in press; Sørensen et al., 2021, submitted; Watson et al., 2020). **(6)** Lower subjective ratings of the conversational success and increased ratings of wanting to improve the situation as well as increased ratings of listening effort (Tuomainen et al., 2019) and talking effort.

5.2 Methods

5.2.1 Participants

For the current study, 11 pairs of native-Danish conversational pairs were recruited. All pairs consisted of an older hearing-impaired (HI) participant (mean age 74.1, sd = 3.5, range 67.8 - 79.1 years), and a younger normal-hearing (NH) participant (mean age 25.3, sd = 6.1, range 19.9 – 39.1 years) who were not previously acquainted. Of the 11 HI participants, six (54.5%) were female, and three (27.3%) of the NH participants were female. The pairs were matched at random, without considering gender, resulting in five pairs of mixed gender, and six pairs of the same gender. The experimental design was originally meant to include 12 pairs, however, due to increased COVID-19 restrictions, the data collection had to be stopped after 11 pairs. All HI participants had symmetrical, mild-to-moderate hearing loss with typical, high-frequency sloping N2/N3 audiograms (Bisgaard et al., 2010). Pure-tone thresholds were determined for the HI participants prior to the experimental visit (Figure 5.1, mean PTA = 35.1, sd = 7.5, range 23.1 – 48.8 dB HL). No significant correlation was found between the age and PTA of the HI participants ($r = -0.17, p = .6$). All HI participants were experienced hearing-aid users and reported using their hearing aids all day (81.8%), or for specific purposes such as watching television, work, or social events (18.2%). The hearing status of the NH participants was assessed by confirming that 20 dB HL tones at 500, 1000, 2000 and 4000 Hz were audible on both ears. All participants gave their written informed consent and the study was approved by the regional ethical committee of the Capital Region of Copenhagen, Denmark (Board of Copenhagen, Denmark, reference H-20068621).

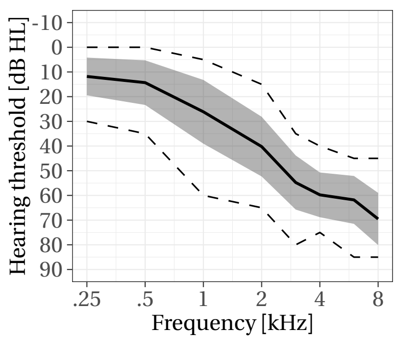

Figure 5.1: Pure-tone hearing thresholds for the hearing-impaired participants averaged across ears. The bold black line indicates the average hearing threshold across subjects and the shaded area the standard deviation. The dotted lines indicate the minimum and maximum hearing thresholds.

5.2.2 Experimental conversational task

Communication between the conversational partners was initiated using the DiapixDK spot-the-difference task (Sørensen, 2021). The pairs were instructed to identify 10 of the 12 differences existing between the two near-identical pictures and to do this as fast as possible within 10 minutes. The pairs were seated face-to-face with 2 meters between them in a soundproof booth. The pairs were asked to solve Diapix tasks in a 2x2 design varying the presence of background noise (quiet and noise) and hearing-aid amplification (see Section 5.2.3) provided to the HI participant (unaided and aided). The pairs were also subjected to a fifth condition with an alternative hearing-aid signal processing scheme in noise. However, due to technical issues, the implemented beamforming was not as narrow as desired, resulting in little change in SNR compared to the omnidirectional condition. The data from the directional condition (i.e., with beamforming) is not included in the following as no significant difference between the omni- and directional experimental conditions were identified. In the conditions with noise, a 20-talker babble (Sørensen et al., 2021, submitted) was presented from two loudspeakers (JAMO D400) positioned between the two participants at an angle of 45- and 315-degrees azimuth, ensuring that both participants experienced the same objective effect of the noise. The background noise was presented at a level of 70 dB SPL, calibrated using a sound-level meter (Brüel & Kjaer, Type 2250) positioned at the expected ear-height of the participants on the empty chairs of the experimental setup.

After the HI participant was fitted with a pair of hearing aids, the two conversational partners were instructed in solving the Diapix task and were given one training run where a single Diapix was solved under the supervision of the test leader. The participants solved two Diapix tasks for each condition, separated into two experimental blocks with a pause in-between. The conditions were counterbalanced across blocks and pairs, while the Diapix images were counterbalanced across conditions and pairs. At the end of the second block, the participants were asked to evaluate their experience in a free conversation performed in quiet. Data from this condition will not be presented here. Before beginning each block, each participant recorded a calibration signal to estimate the speech level (see Section 5.2.4).

Inspired by Picou and Ricketts (2018), the participants were asked to answer four questions after finishing each Diapix, relating to the communication during the task-solving. The questions were (translated from Danish) 1) How successful do you think the conversation was? 2) If this situation occurred in your everyday life, how likely would it be that you would try and improve the situation (e.g., by moving to a different room, ask your partner to speak louder)? 3) How effortful was it to speak? 4) How effortful was it to listen? All questions were answered by making a mark on a 0-10 visual-analog scale ranging from “Not at all” to “A lot”.

5.2.3 Hearing-aid fitting

The HI participants were fitted with Signia Pure 312x RIC device with a closed standard dome (click-sleeve), with the gain determined by the NAL-NL2 rationale (Keidser et al., 2012). Closed fittings were chosen with the expectation that applying directionality to a closed fitting would yield a greater effect. Unfortunately, closed-fittings can result in altered own-voice perception due to a lack of direct sound inputs. For this reason, the Own-Voice Processing technology in the Signia hearing aids was enabled (Powers et al., 2018). The Own-Voice Processing feature is trained in the hearing-aid fitting procedure to detect the wearer’s own voice based on spectral content and the direction of arrival, making it robust to changes in the speech, e.g., resulting from Lombard effects. Upon detecting the wearer’s own voice, the gain provided by the hearing aid is reduced as to obtain a more natural perception of their own voice (Høydal Harry, 2017).

A static omni-directional program was made by disabling digital noise reduction, speech enhancement features (SpeechFocus), transient noise reduction (SoundSmoothing), and the directional pattern approximating the pinnae effect (TruEar) (Beilin and Powers, 2013). As various hearing-aid parameters were logged throughout the experiment (data not presented here), a designated program was made for the unaided condition where the gain in all frequency bands was dampened as much as possible (from 0 dB for low frequencies up till 8 dB for high frequencies). In the unaided experimental condition, the test leader changed to the dampened program and removed the receivers from the ears, but left the hearing aid hanging on the ears of the HI participants.

5.2.4 Sound recordings and analysis

Each participant was equipped with a cheek-mounted directional microphones (DPA 4088) connected to a Mackie 402 VLZ4 pre-amplifier and an ECHO Audiofire12 soundcard through a Neutrik patchbay (NYS-SPP-L 48). All sounds were recorded at a sampling frequency of 44100 Hz, using Matlab 2018a, and saved in wav-format for offline processing. In order to estimate the speech level of each participant, a calibration measurement was performed at the beginning of each experiment block. In turn, each participant was seated in their chair and asked to introduce themselves, while a 5-second speech signal was recorded from the cheek-mounted microphone and an omni-directional reference microphone (Behringer B-5) placed on the empty chair where the conversational partner would sit and adjusted to the expected height of the partner's ears. With these two signals, the attenuation, in RMS, from the cheek-mounted microphone to the microphone placed at the chair of the conversational partner could be calculated. To convert this into dB SPL, a calibration signal recorded prior to the experimental visit was used. The calibration signal was recorded from the reference microphone placed on the empty chair in close proximity to a sound-level meter (Brüel & Kjaer, Type 2250). A white noise signal was then presented from one of the two loudspeakers at a level of 75 dB SPL, confirmed by the sound-level meter, and recorded by the reference microphone. For each experimental condition, the speech level at the position of the conversational partner could then be estimated by subtracting the attenuation from the cheek to the reference microphone and converting it into dB SPL by normalizing by the RMS level to that of the calibration signal.

Offline, from the recordings of the Diapix task obtained from each cheek-microphone, power-based Voice Activity Detection (VAD) was used to identify and categorize individual utterances, following the approach in Heldner and Edlund (2010), Sørensen et al. (in press), and Sørensen et al. (2021, submitted). In each 5 ms window (with 1 ms overlap), the segment was labeled as containing speech if it was above an individually set power threshold. Speech intervals with gaps shorter than 180 ms were merged, and intervals shorter than 90 ms were removed to avoid categorizing transient sound bursts as speech. The resulting utterance detections for each pair and condition were then fed to a communicative state classification algorithm (Sørensen et al., in press), labeling utterances into the following categories: overlaps-between (acoustic overlaps between the turns of interlocutors during a floor transfer), gaps (acoustic gaps between the turns of interlocutors during a floor transfer), pauses (pauses in one person's speech stream that did not result in a floor-transfer), overlaps-within (stretches of speech that occured completely within the other interlocutor's turn), and interpausal units (IPUs; stretches of speech surrounded by 180 ms of silence, not including overlaps-within). Together, overlaps-between and gaps make up floor-transfer offsets FTOs, where negative FTOs are overlaps-between and positive FTOs are gaps. Further, we calculated the speaking time of both participants in the conversation, and based on the average RMS of each utterance and the calibration offset, the speech level of each participant was calculated. Finally, a Praat script for detecting syllables (De Jong and Wempe, 2009) was used with a silence threshold of –25 dB, a minimum dip between peaks of 2 dB, and a minimum pause duration of 180 ms. In an interface between Praat and MATLAB, we extracted the detected syllables in the utterances we had detected with the VAD and normalized by the duration of those utterances to compute the articulation rate for each person in each condition and block.

5.2.5 Statistical analysis

The *lme4* package for *R* (Bates et al., 2015) was used to build mixed-effects regression models for each of the variables of interest. Unless otherwise stated, the maximal starting model before reduction included fixed effects of processing (unaided, aided), background (quiet, noise), hearing (normal, impaired), and block (1, 2) with up to second-order interactions, as well as a random intercept of pair and person varying within pair, i.e.: $x \sim$ processing + background + hearing + block + processing:background + processing:hearing + processing:block + background:hearing + background:block + hearing:block + (1 | pair/person). The backward stepwise elimination of non-significant factors (*step* function in the *lmerTest* package, Kuznetsova et al., 2017) was used to reduce the models with an alpha level set to 0.1 to avoid stepping out borderline significant factors. The *anova* function from the *stats* package and residuals plots were used to compare models before and after reduction to ensure the one with the better fit was selected. Q-Q plots and the Anderson-Darling test for normality were used to confirm that the residuals of the final model were normally distributed, which was the case for all variables. Any post-hoc analyses were done using the *ls_means* function from the *lmerTest* package, computing pairwise differences of least-squares means using a Satterthwaite method for estimating the degrees of freedom.

5.3 Results

5.3.1 Task completion and speaking time

The time it took the pairs to identify 10 differences between the Diapix can be used as an indicator of the efficiency of the conversation (Baker and Hazan, 2011). For time considerations, the pairs were stopped after 10 minutes if they had not yet found 10 differences. This occurred in three cases (1.1% of all trials, two in quiet and one in noise), and the completion times for these cases were not included in the analysis. The task completion time increased by an average of 45 s for conditions with background noise [$F(1,148.9) = 20.5, p < .001$], see Figure 5.2. A significant learning effect was observed between the two blocks [$F(147.9) = 8.0, p < .01$], such that the pairs solved the task 28.4 s faster in the second block.

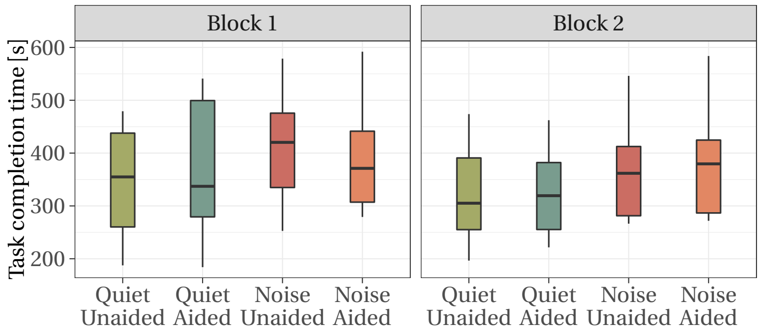

Figure 5.2: The completion times across conditions and blocks extracted for each pair. Here, and in the following plots, the boxes indicate the 25th to 75th percentile and the horizontal line the median. The whiskers extend the range of the data, and outliers will be indicated with dots. The two panels share a common y-axis.

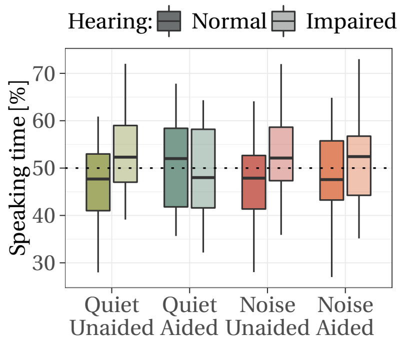

Figure 5.3: Percentage speaking time between the two interlocutors of each pair. A completely balanced partaking of the conversation of 50% is indicated by the dotted line.

From the detected utterances, the relative speaking time of each participant was calculated. As the sum of the relative speaking time for each pair is always 100%, the statistical analysis was performed on data from the HI participants only. The analysis showed an effect trending towards significance of processing [$F(1,76) = 3.4, p = .068$]. In a post-hoc analysis, it was revealed that there was only a significant decrease in speaking time when HI participants were aided in quiet [$t(74) = 2.25, p < .05$], but not in noise [$t(74) = .397, p = .69$]. Overall, this suggests that HI interlocutors contribute less to the conversation in quiet when aided compared to the other conditions.

5.3.2 Subjective ratings of the conversations

The subjective ratings made after each Diapix task (see Figure 5.4) were all affected by the presence of background noise, resulting in: 1) the conversational success [$F(1,152.09) = 98.2, p < .001$] being rated 2 points lower, 2) the desire to improve the situation [$F(1,151.1) = 656.9, p < .001$] being rated 6.6 points higher, 3) the talking effort [$F(1,162.03) = 421.3, p < .001$] being rated 5.3 points higher, and 4) the listening effort [$F(1,151.1) = 430.3, p < .001$] being rated 4.9 points higher.

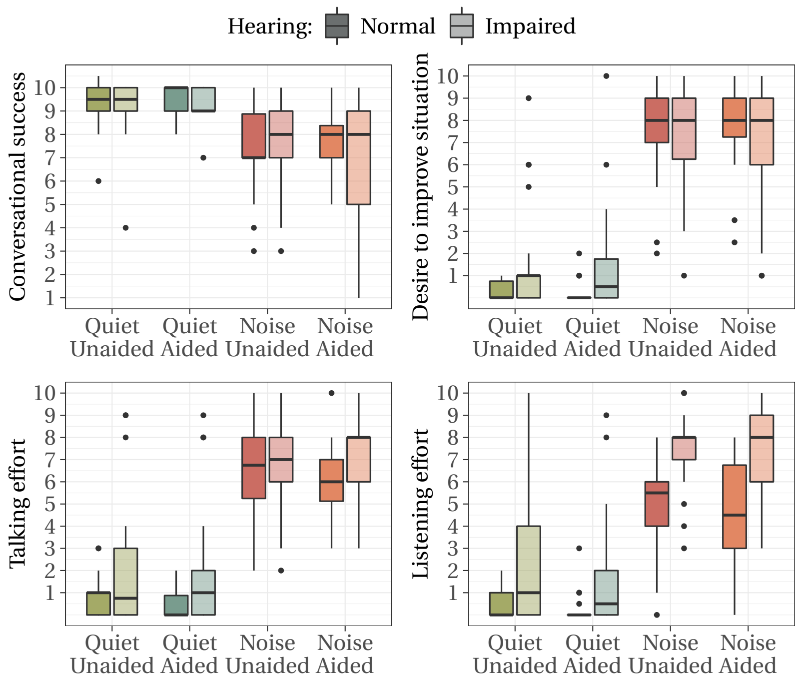

Figure 5.4: Subjective ratings of the conversation. On a scale from 0 – 10, each subject independently rated the conversational success (upper left), the desire to improve the situation (upper right), as well as the effort related to talking (lower left) and listening (lower right).

The HI participants generally rated their talking [$F(1,162) = 16.9, p < .001$] and listening effort [$F(1,20) = 11.1, p < .01$] higher than their NH partner. The listening effort of the HI participants was also affected by the background noise [$F(1,151.1) = 5.4, p < .05$], such that in quiet the HI participants rated their listening effort as being 1.6 points greater than the NH participants [$t(25.5) = -2.4, p < .05$], while it was rated 2.7 higher in noise [$t(25.5) = -4.0, p < .001$]. The background noise also affected the HI participants' desire to improve the situation [$F(1, 151.1) = 12.2, p < .001$]. Considering that the HI participants rated higher listening and talking effort and a higher desire to improve the quiet situation, it is interesting to observe that they rated the conversational success the same as their NH conversational partners. One common criterion on which conversational success could be judged is the task completion time. To investigate this and whether the rating of conversational success was linked to any of the other subjective ratings, a model predicting the conversational success from the three other subjective ratings and completion time was built (no interaction effects included). The result showed that the rating of conversational success was positively related to the completion time [estimated coefficient = -0.004, $F(1.139.2) = 8.7, p < .01$], and the desire to improve the situation [estimated coefficient = 0.1, $F(1, 162.4) = 4.0, p < .05$], and negatively related to the talking [estimated coefficient = -0.25, $F(1, 162.9) = 15.9, p < .001$] and listening effort [estimated coefficient = -0.2, $F(1, 135.4) = 9.9, p < .01$].

5.3.3 Floor-transfer offsets

From the recorded conversations, the FTO was extracted for each floor transfer. The FTO distributions, left panel of Figure 5.5, peak on average around 230 ms, indicating that participants tend to initiate a turn with a small gap, but overlaps and longer gaps between talkers occur too. From a visual inspection, the HI participants' distributions (dotted lines) seem broader than that of their NH interlocutor, indicating decreased precision in turn-timing, but in quiet when the HI is aided, the NH and HI look very similar. For each participant, block, and condition, the median and interquartile range (IQR) was extracted from the distribution to evaluate the effects of the experimental contrasts on the timing of turn-taking.

The median FTO (right panel of Figure 5.6) was affected by background noise [$F(1, 151) = 45.2, p < .001$], causing the participants to initiate their turn 69 ms later in noise [$t(151) = 6.7, p < .001$]. A near-significant interaction between hearing status and HA processing [$F(1, 151) = 3.9, p = .051$] indicates that when aided, the HI participants started their turns 41 ms earlier [$t(1, 151) = -1.97, p =$

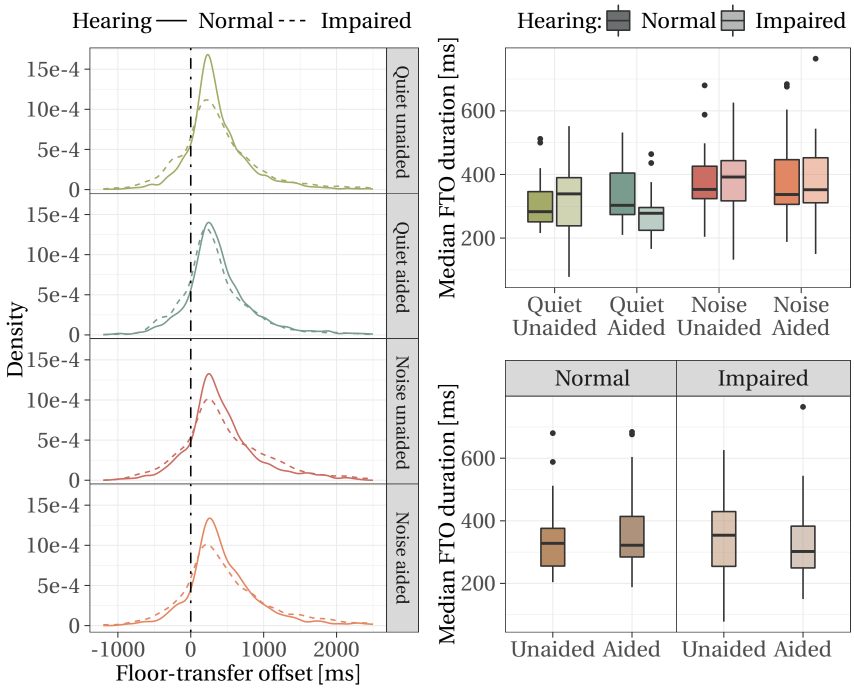

Figure 5.5: Distribution of FTOs and median FTOs. Left panel: The distributions of FTOs pooled across participants and block for normal-hearing (solid lines) and hearing-impaired (dotted lines) in each condition. Positive FTO values indicate a gap between talkers at the point of turn-taking, while a negative FTO indicates an overlap between them. Right panel: The median extracted from the FTO distribution of each participant and condition, averaged across blocks (upper right) and averaged across background and block (lower right) to visualize interaction effects.

.051].

The variability of the turn-taking timing (FTO distribution), as measured by the IQR, was significantly larger in noise [$F(1,148) = 59.4, p < .001$] for both groups (see upper left panel of Figure 5.6), and the IQR of the HI participants was larger than that of the NH participants [$F(1,20) = 13.1, p < .01$], see upper right panel of Figure 5.6. Furthermore, the FTO variability was affected by a total of three interaction effects. First, the variability in turn taking varied between background conditions for the NH and HI [$F(1,148) = 8.4, p < .01$], see lower left panel of Figure 5.6. In quiet, the difference between NH and HI was 92 ms [$t(28.3 = -2.1, p < .05$], while it was 193 ms in noise [$t(28.3) = -4.5, p < .001$]. The NH had an increase in variability of 83 ms in noise compared to quiet [t(148) = -3.4, p < .001] whereas it was 184 ms for the HI [$t(148) = -7.5, p < .001$].

Second, hearing status and HA processing also resulted in a near-significant

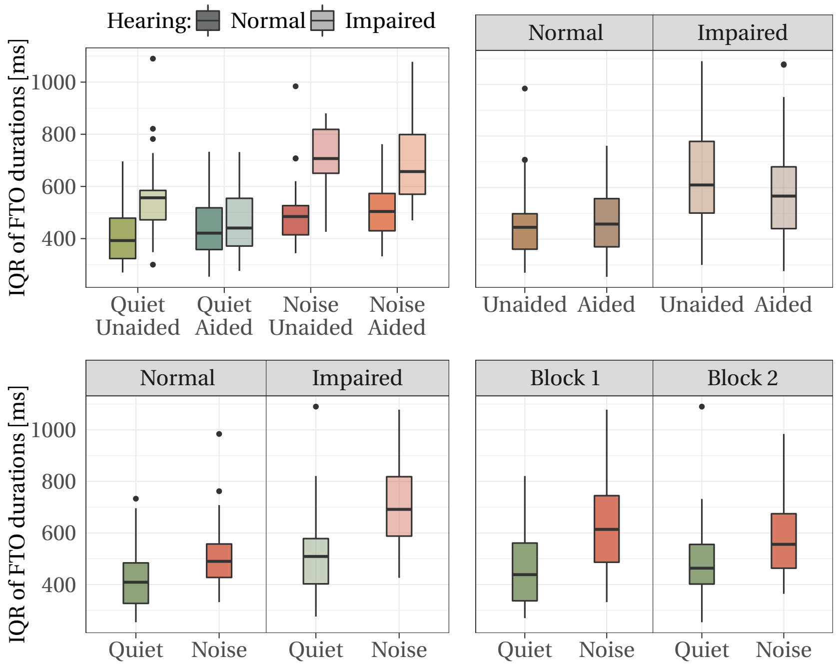

Figure 5.6: The interquartile range (IQR) extracted from the FTO distribution of each participant and condition averaged across blocks (upper left), for the NH and HI averaged across backgrounds (upper right), for the NH and HI averaged across processing (lower left), and for block 1 and block 2 averaged across processing and hearing (lower right) to visualize interaction effects. The four panels share a common y-axis.

interaction $[F(1,148) = 3.7, p = .058]$, suggesting that when aided, the variability of the HI participants decreased by 50 ms compared to when they were unaided $[t(148) = 2.0, p < .05]$, see upper right panel of Figure 5.6.

Finally, an interaction between experimental block and background noise was identified $[F(1,148) = 4.0, p < .05]$, see lower right panel of Figure 5.6. The post-hoc analysis revealed a larger difference of 160 ms in the variability between quiet and noise in the first block $[t(148) = -6.8, p < .001]$, while the difference was only 99 ms in the second experimental block $[t(148) = -4.1, p < .001]$.

Figure 5.7 shows the rate of floor transfers (FT) per minute. The FT rate decreased in the presence of background noise $[F(1,162) = 45.6, p < .001]$, see left panel of Figure 5.7, and there was a significant effect of block $[F(1,162) = 5.4, p < .05]$. Further, there was an interaction effect between background noise and experimental block $[F(1,162) = 9.7, p < .01]$, see right panel of Figure 5.7, indicating that while the noise caused interlocutors to decrease their FT rate

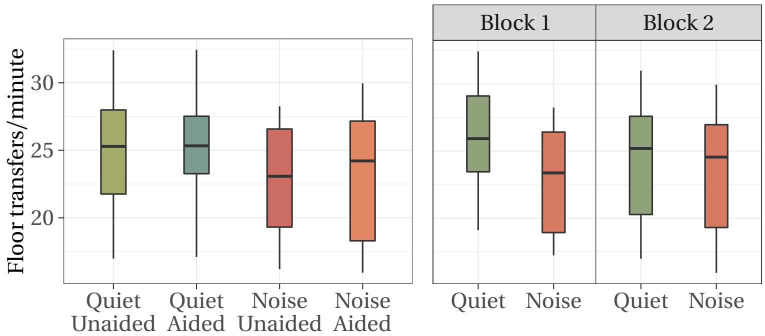

Figure 5.7: The number of floor transfers per minute between the two talkers averaged across block (left panel) and averaged across processing and hearing to visualize interaction effects (right panel). The two panels share a common y-axis.

by 2.9 per minute in the first experimental block [$t(162) = 6.9, p < .05$], the reduction in FT rate in noise relative to quiet was only 1.1 per minute in the second block [$t(162) = 2.5, p < .05$]. This effect stems mainly from the fact the number of FTs in quiet was reduced by 1.6 FTs per minute between the first and second block [$t(162) = 3.8, p < .001$], whereas the number of FTs in noise did not change between blocks [$t(162) = 0.55, p = .58$].

5.3.4 Speaking pace, level, and duration

In Figure 5.8, the speech levels estimated at the interlocutor's position in dB SPL are shown. In noise, the talkers increased their speech levels significantly [$F(1,149) = 2310, p < .001$], see left panel of Figure 5.8. The HI participants spoke at a slightly negative SNR of -0.7 dB (median value), while the NH participants spoke at a slightly positive SNR of 1.1 dB. The background noise also affected the speech levels in interaction with processing [$F(1,149) = 5.4, p < .05$], see middle panel of Figure 5.8, with the post-hoc analysis revealing that when the HI participant was unaided, they spoke 1.1 dB louder in quiet [$t(149) = 3.0, p < .01$], whereas the hearing-aid amplification did not affect the speech levels produced in noise [$t(149) = -0.26, p = .78$]. A significant interaction between hearing status and background noise was also identified [$F(1,149) = 31.2, p < .001$], see right panel of Figure 5.8, showing that when adding background noise, the NH participants raised their speech levels by 13.6 dB [$t(149) = 18.5, p < .001$], whereas the HI participants only raised their speech levels by 10.8 dB [$t(149) = -29.9, p < .001$]. There was a significant effect of noise on the articulation rate,

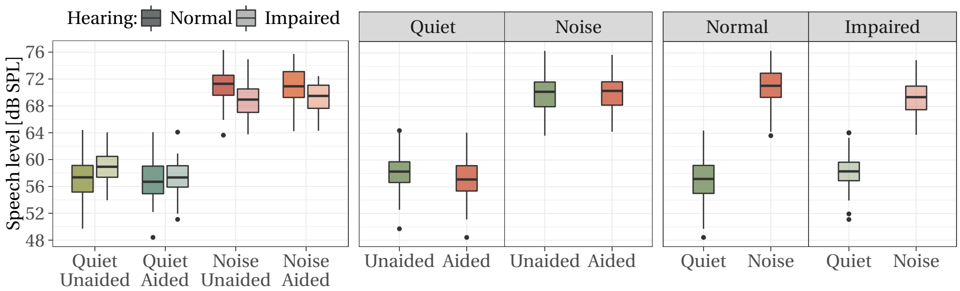

Figure 5.8: Estimated speech levels at the average position of their conversational partner in dB SPL, averaged across block (left panel), averaged across block and hearing status (middle panel), and averaged across block and processing (right panel) to visualize interaction effects. The three panels share a common y-axis.

see left panel of Figure 5.9, $[F(1,149) = 11.8, p < .001]$. However, an interaction effect between background noise and experimental block $[F(1,149) = 7.9, p < .01]$ revealed that while participants spoke slower in noise compared to quiet in block 1 $[t(149) = 4.4, p < .001]$, there was no significant difference between the articulation rates in noise and quiet in the second block $[t(149) = .45, p = .65]$, see middle panel of Figure 5.9. This was caused by interlocutors speaking significantly slower in quiet in the second block $[t(149) = 2.22, p < .05]$. The articulation rate was also affected by a near-significant interaction between hearing status and HA processing $[F(1,149) = 3.8, p = .053]$ indicating that the HI participants spoke faster when aided $[t(149) = -2.4, p < .05]$. Hearing-aid processing did not affect the NH participants' articulation rates $[t(149) = 0.31, p = .76]$, see right panel of Figure 5.9. The median IPU duration, i.e., the

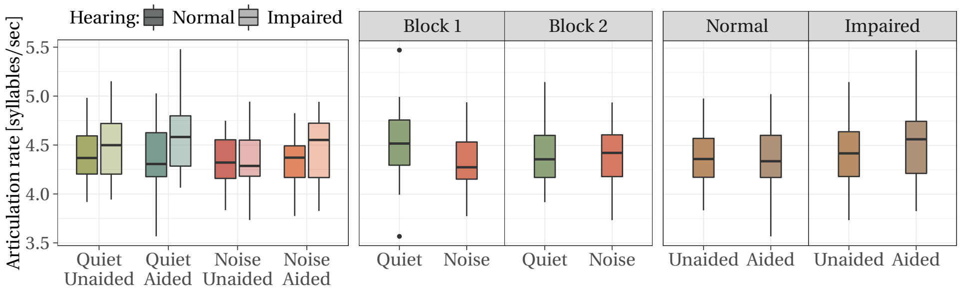

Figure 5.9: The articulation rate measured as the number of syllables per second of speaking time per person averaged across block (left panel), averaged across background and block (middle panel), and averaged across hearing and processing (right panel) to visualize interaction effects. The three panels share a common y-axis.

median lengths of connected speech surrounded by 180 ms of silence, is seen in the left panel of Figure 5.10. Background noise $[F(1,149) = 9.5, p < .01]$

caused participants to increase their IPU durations [$t(149) = -3.1, p < .01$]. The median IPU was also affected by an interaction between background noise and experimental block [$F(1,149) = 5.2, p < .05$], see right panel of Figure 5.10. The post-hoc analysis revealed that while there was no difference in the median IPU durations in noise between blocks [$t(149) = .8, p = .4$], participants held their turn for 58 ms longer in quiet in block 2 compared to block 1 [$t(149) = -2.4, p < .05$], see middle panel of Figure 5.10. An interaction between hearing and processing [$F(1,149) = 5.5, p < .05$] revealed that when unaided, the HI participants talked for 134 ms longer than their NH conversational partners [$t(149) = -2.3, p < .05$], while there was no difference between the two talkers when the HI participant was aided [$t(149) = -0.9, p = .36$], see right panel of Figure 5.10.

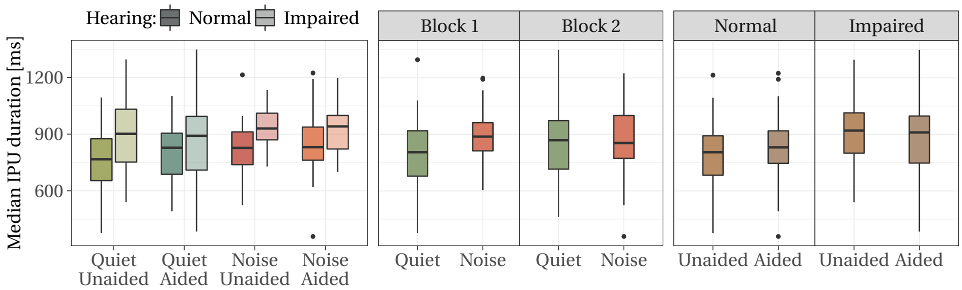

Figure 5.10: Median IPU durations averaged across block (left panel), and averaged across background and block (middle panel panel), and across hearing and processing (right panel) to visualize interaction effects. The three panels share a common y-axis.

5.4 Discussion

By exploring the dynamics of a conversation between a younger NH and an older HI talker, the goal of the current study was to investigate the effects of adding background noise and compensating for the HI participants' reduced audibility by providing them with amplification in hearing aids. Our findings can briefly be summarized as follows: 1) The addition of background noise resulted in large changes in conversational behavior. 2) Providing amplification to the HI participants alleviated some of their communication difficulties. 3) Hearing impairment had detrimental effects on communication, especially in background noise. 4) The pairs adapted their conversation strategy over time. In the following, each of the four points will be discussed in more detail. Finally, a cross-study comparison will be provided to address the potential differences of being seated face-to-face compared to sitting in different rooms.

5.4.1 Background noise impacts multiple aspects of the conversations

As hypothesized, the presence of background noise caused alterations in the communication evident from longer completion times (Figure 5.2), louder speech, slower articulation, and longer IPUs (Figure 5.8-Figure 5.10), as well as fewer, later, and less well-timed turn starts (Figure 5.6 and Figure 5.7). In noise, the participants rated the conversations as being less successful, they had an increased desire to improve the situation, and they experienced more talking and listening effort (Figure 5.4).

When the 70 dB SPL babble background noise was presented, the participants raised their speech levels such that the NH participants spoke at an average positive SNR of 1.1 dB, while the HI participants spoke at an average negative SNR of –0.7 dB (Figure 5.8). Both in the current study and the study by Sørensen et al. (2021, submitted), communication was possible at SNRs close to 0 dB. In comparison, in most open-set speech-in-noise tests, HI listeners often require SNRs above +5 dB SNR to correctly repeat only 50% of a target sentence (Wilson et al., 2007). This indeed highlights that communication taps into different resources and processes than merely listening and repeating sentences; when there is context and the possibility to do error correction, and people are adapting to each other's difficulty, interlocutors may not require as high an SNR to maintain a fluid conversation.

Despite participants in the current study adopting their speech production to improve the communication situation, changes in the timing of turn-takings were still observed (Figure 5.6). Indeed, background noise resulted in later (larger FTO median, upper right panel of Figure 5.5) and more variable turn starts (larger FTO IQR, upper left panel of Figure 5.6). Achieving rapid turn starts observed in regular conversations requires correctly predicting the end of one's interlocutor's turn and preparing a verbal response (Corps et al., 2018b), both of which rely on being able to hear and interpret the interlocutor's speech. In suboptimal listening conditions, the model for ELU describes how explicit processing is activated, involving, e.g., inference-making and semantic integration, which requires longer processing time to obtain speech understanding (Rönnberg et al., 2013). In this context, producing longer utterances could give a conversational partner more time to understand and prepare an adequate response. Indeed, Sørensen et al. (in press) found that longer IPUs were proceeded by faster and less variable FTOs, and Beechey et al. (2018), Sørensen et al. (in press), Sørensen et al. (2021, submitted), and Watson et al. (2020) all found that interlocutors increased their IPU durations in noise. Sørensen et al. (2021, submitted) found that HI produced even longer utterances than their NH interlocutor, as we found in the current study when HI participants were unaided. As was speculated in Sørensen et al. (2021, submitted), if people experience a degraded input signal and thus have had to spend additional resources planning their response, they may not have planned the entirety of their response before launching it at the socially appropriate time. Thus, they may continue planning their utterance while producing it, which would require a longer time, which in return could lengthen the IPU durations. They could insert filler-words such as “uhm” (Clark and Fox Tree, 2002; Sjerps and Meyer, 2015) at the boundary and/or during the turn to buy time to prepare a complete response, which should lengthen their utterances.

Considering that participants speak for longer (IPU median) and the timing of their turn starts are later (FTO median) and more variable (FTO IQR), it is not surprising to find that the FT rate is slower in noise (Figure 5.7). Assuming that the same amount of information is conveyed between the conversational partners in noise and quiet, this will consequently lead to increased completion times, as observed in this study (Figure 5.2). However, an exploratory analysis showed that the influence of FT rate on completion time was not significant [$p = .09$], and the estimated coefficients revealed that a 1% decrease in FT rate only caused a 0.2% increase in the completion time. A similar analysis for articulation rate also showed that slowing down the speech did not significantly influence the completion time [$p = .23$]. Hence, it is likely that the prolonged completion times were more affected by miscommunications or alterations in the linguistic information conveyed between interlocutors.

The various detrimental effects that noise has on the communication efficiency is experienced by the participants, who all subjectively rated the conversational success to be lower, while they rated their desire to improve the situation as well as their talking and listening effort higher. The increase in rated talking and listening effort was accompanied by changes in the speech levels (Figure 5.8). The subjective rating of conversational success was included to get a rating of the overall conversation, and the statistical analysis confirmed that the three remaining ratings were significant predictors of conversational success together with the completion times.

5.4.2 Hearing-impaired participants experienced increased communication effort, especially in noise

The communication behavior differed significantly between the two conversational partners based on their hearing status; however, differences were most eminent in the presence of background noise. The most noticeable difference between the participants was observed in the change of speech level between noise and quiet (right panel of Figure 5.8). Here, the HI participant increased their speech level by 10.8 dB, whereas the NH participants had to increase by a further 2.9 dB to ensure efficient communication. The fact that the NH participant had to increase their speech level in noise further when the speech levels did not differ in quiet underlines the known fact that communicating in noisy situations is especially difficult for HI people.

Hearing impairment also affected the ability to accurately time turn starts, as the FTO variability was 142 ms higher for the HI participants (Figure 5.6). Interestingly, the median FTO was not affected by hearing status (Figure 5.5); however, the HI participants were much less consistent in the timing of their turn starts (FTO IQR, Figure 5.6). This might suggest that the HI participants spent additional cognitive resources on processing the speech, potentially resulting in fewer resources left for the processing load associated with planning an upcoming response (Barthel and Sauppe, 2019). Had hearing impairment only affected the ability to process and understand speech, we would have expected the median FTO, but not IQR, to increase. However, a more variable FTO indicates that the HI participants' ability to predict the end-of-turn timing of their partner is reduced. This ability is further impaired for the HI participants in the presence of background noise, highlighting how detrimental background noise is to the communication of HI listeners.

Indeed, the HI participants also indicate this increased communication difficulty in the subjective ratings, reporting an increased desire to improve the quiet situations and having a significantly higher talking and listening effort. While adding background noise makes the HI participants report an additional increase in listening effort than their NH conversational partners, no such interaction effect was observed for the talking effort.

It has been previously noted that a participant can subjectively interpret rating listening effort to include the levels of chronic stress or fatigue (McGarrigle et al., 2014; Pichora-Fuller et al., 2016), which are known to be elevated in HI listeners (Hornsby et al., 2016; Nachtegaal et al., 2009). The current finding that listening, but not talking effort, was elevated in HI listeners in quiet and more so in noise serves as an indication that the participants were able to disentangle the evaluation of listening and talking from other internal interpretations of effort. Despite this, it is interesting that although the HI participants produced less well-timed turn starts (increased FTO IQR), this did not affect the subjective rating of the talking effort. On a similar note, the speech levels were higher for the HI in quiet, while the NH increased their levels more in the presence of noise, none of which were evident from their subjective ratings of talking effort. It was investigated whether the listening and talking effort ratings were affected by the speech level produced by the talker and conversational partner, respectively, but no statistical relationship was found (and the analysis has been omitted from the Results section). Hence, it is unknown what internal criterion drives the subjective ratings of listening and talking effort in the current study.

The participants were also asked to rate the conversational success to investigate if they subjectively experienced alterations in the communication. As the HI participants rated experiencing higher talking and listening effort, as well as a higher desire to want to improve the situation, it was surprising that no effect of hearing status was found on the rating of conversational success. The explanation for this is most likely that the completion time served as a common internal criterion by which the conversational success was judged. It was observed that the individual rating of conversational success was significantly related to the ratings of the desire to improve the situation and the talking and listening effort. However, we must conclude that conversational success is not a well-understood concept; indeed, in the scientific literature, conversational success has been evaluated as the lack of conversational breakdown (Beechey et al., 2020b; McInerney and Walden, 2013), but also as depending on the conversational topic, social contact, fluency and other factors (Lind, 2012). As such, it is potentially more beneficial to ask participants to rate more well-defined subsets of a conversation than the overarching success of it.

5.4.3 Participants adapt over time

The experimental Diapix task was used to ensure a flowing conversation between the two previously unacquainted participants. However, in line with previous studies (Sørensen et al., 2021, submitted), we also observed a significant learning effect over time, resulting in tasks in the second experimental block being solved 30 seconds faster than in the initial block (Figure 5.2). One reason for this is of course the familiarity of the task; developing a task-solving strategy, learning the detail-level of the differences in the pictures and where they most often occur. Besides task-related learning effects, alterations over time can also occur because of alterations in the social familiarity between the participants.

The articulation rate was observed to slow down between quiet and noise in the first block, whereas it did not in the second, primarily caused by a lower articulation rate during the quiet conditions of the second block. This is a puzzling finding considering that slowing down speech is interpreted as a sign of hyper-articulation associated with wanting to improve difficult communication situations (Hazan and Baker, 2011). We cannot explain why participants slowed their articulation rate in the easiest condition (quiet) when being well-acquainted with the experimental task (block 2).

For the turn-taking measures, both the variability (FTO IQR, lower right panel of Figure 5.6) and rate (FT rate, right panel of Figure 5.7) were less affected by noise in the second experimental block. The latter is primarily driven by a reduction in the number of floor transfers in the quiet condition of block 2, as the number of floor transfers in noise did not change between blocks. This might suggest that the interlocutors are more familiar with each other's speaking patterns and can time their turn-taking better (lower FTO IQR) and require fewer floor transfers to identify the Diapix differences.

5.4.4 Hearing-aid amplification affects the conversation

As hypothesized, providing the HI participants with amplification resulted in fewer and more moderate effects than adding background noise. Providing amplification resulted in a more balanced speaking time between the conversational partners in quiet (Figure 5.3), reduced speech level of both talkers in quiet (middle panel of Figure 5.8), while the HI participants increased their articulation rate (right panel of Figure 5.9), reduced the IPU duration to match that of the NH participants (right panel of Figure 5.10), and showed ealier and less variable timing of turn starts (upper right panel of Figure 5.6) when provided with hearing-aid amplification.

Although speaking time is not a measure of communication success, it is believed that an equal distribution of speaking time indicates a more cooperative conversation (Beechey et al., 2019). The balance is of interest in this study, as HI talkers are known to adapt the face-saving strategy of dominating a conversation in order to avoid listening (Jaworski and Stephens, 1998; Stephens and Zhao, 1996). Indeed, evidence for such a strategy has been observed when participants are not facing each other, resulting in the HI talkers contributing, on average, with 57% of the conversation in 70 dBA SPL noise (Sørensen et al., 2021, submitted). In the current study, HI participants did not dominate the conversation, however, HA amplification did cause the HI participants to contribute less in quiet (Figure 5.3), causing a more balanced contribution between the conversational partners. This observation suggests that when audibility is restored, the HI talker will be more inclined to listen, rather than talk. This is also evident from the observed reduction in median IPU durations for the HI participants when aided to a length not significantly different from that produced by the NH talkers.

Amplification also affected the timing of turn starts (FTO median and IQR), making the HI participants 41 ms faster in initiating their turn and reducing the variability by 50 ms. As argued previously, the timing of turn starts could rely on the cognitive processing of the speech input, suggesting that hearing-aid amplification improves the HI participants' ability both to predict the end-of-turn and prepare the verbal response. This is in sharp contrast to standardized listening tests, where no social pressure is put on the participants to formulate a well-timed response.

The reduced speech levels observed for both participants when the HI talker was aided, suggests that the loss of hearing modifies the speech production by causing a Lombard reflex (Junqua, 1996). In quiet, amplification makes up for the lost audibility of the HI participant's own voice and improves the speech intelligibility of their interlocutor, causing a reduction in the speech level of the HI participants and consequently of their NH interlocutor.

The phenomena *ampclusion* (Painton, 1993) describes how the hearing-aid wearers' perception of their own voice is changed, not only by the amplification thereof but also by altered auditory feedback arising from the closed-dome hearing aids occluding the ear canal. A recent study found that experienced hearing-aid users rated their voice as being more dominating after receiving new hearing aids, in fact even more so than a group of participants receiving their first hearing aids (Hengen et al., 2020). If the hearing aids provided in the current study made the HI participants feel that their voice was being too dominating, we would expect the speech level alterations between unaided and aided of the HI participants to be similar in noise and quiet, whereas we only observe changes in the former. Hence, we do not suspect that altered own-voice processing to be the main driving force of the observed effects of amplification.

In summary, providing hearing-aid amplification causes the HI participants to alter their communication and become more similar to their NH conversational partner in quiet. Significant effects of hearing-aid amplification in quiet for listeners with mild to moderate hearing impairment are not easily observed in most traditional psychoacoustic tests. To quantify the benefit of hearing-aid processing in close-to-ideal listening situations, studies often investigate secondary cognitive effects such as delayed recall of target words (Kuk et al., 2021; Lunner et al., 2016), or focus on changes in biological markers of listening effort, such as pupillometry (Wendt et al., 2017). In this study, we were able to show the direct effects of amplification on a communication task performed in quiet.

5.4.5 A cross-study comparison of the effect of face-to-face communication

The experimental paradigm of the current study is adapted from that of (Sørensen et al., 2021, submitted), applying the same Diapix task, background noise, and inclusion criteria for the NH and HI participants, but with participants seated in different rooms. By comparing the unaided conditions between the two studies, we can reflect on the differences between communicating face-to-face and when seated in separate rooms, from here on denoted *remote communication*. It should be noted that in the current study the background noise was presented at 70 dBZ SPL (unweighted) for the 11 pairs included, whereas (Sørensen et al., 2021, submitted) presented the noise at 70 dBA SPL directly into the headphones worn by the 24 participants (12 pairs).

In Sørensen et al. (2021, submitted), both the NH and HI spoke at positive SNRs: NH participants spoke at 1.7 dB and HI participants spoke at 1.1 dB SNR, on average (HI participants were unaided). In the current study, NH participants spoke at 1.1 dB SNR, whereas HI participants spoke at -0.7 dB SNR. Although this suggests that face-to-face communication allows people to speak at slightly poorer SNRs, it is surprising that the conversation was not affected more by the access to body language and gesticulation offered in face-to-face communication. A potential source for improving communication is the lip movements accompanied when seated face-to-face (Fitzpatrick et al., 2015), which are exaggerated when speech is produced in the presence of noise (Cooke et al., 2014). A potential reason for the lower SNRs observed in this study could be that Sørensen et al. (2021, submitted) calibrated the headphone-presented levels as if the interlocutors were seated one meter apart and could not alter the SNR by moving closer. In the current study, the speech levels were calibrated to the expected average position of the interlocutor (see Section 5.2.4). However, the participants were able to alter the SNR by leaning forth/back and turn their better ear towards their interlocutor. Recent studies describe, however, how interlocutors, while sitting 1.5 m apart, only decreased their distance by up to 10 cm in noise, resulting in a negligible SNR improvement of less than 1 dB (Hadley et al., 2019). Furthermore, interlocutors tend to turn their heads to favor their better ear, at the expense of reducing the SNR and the ability to do lip reading (Brimijoin et al., 2012).

Due to COVID-19 restrictions, the interlocutors of the current study were placed at a distance of two meters from each other, which is a doubling of the distance compared to that imposed on the remote communication study (Sørensen et al., 2021, submitted). Hence, in the face-to-face communication, the talkers would have had to produce a level that was approximately twice as loud to achieve the same SNR as the interlocutors produced during remote communication, which could be physically very strenuous. This poses another interesting question: whether interlocutors could have raised their levels to achieve even better SNRs in remote communication, but avoided it to minimize the energy spent on communication or for social reasons.

Common for the FTO median, FTO IQR, FTO rate, articulation rate, and IPU duration is that all are less affected when adding noise in face-to-face communication, compared to remote communication. The faster and less varying turn-taking observed in face-to-face communication is in line with the above observation; that visual information improves communication, including the turn-taking timing, potentially through the facilitation of hand gestures (Bekke et al., 2020) and eye contact (Kiessling et al., 2003). The FT rate decreased more in noise in remote, compared to face-to-face communication, potentially driven by the fact that in noise the HI participants increased their IPUs length by 18% in remote communication, but only 4% in face-to-face communication. This increased length of turn-holding for the HI participants in noise for remote communication also influenced the percentage of speaking time, with a greater contribution from the HI in remote communication compared to face-to-face.

It is interesting to note that despite finding differences in the communication difficulty between face-to-face and remote communication, the task completion times in general, and the change occurring when adding background noise, was not affected by face-to-face communication even though in the current study interlocutors were instructed to solve the task as fast as possible, while in Sørensen et al. (2021, submitted) no such instruction was given.

In both face-to-face and remote communication, effects of hearing status were observed. However, the study designs do not allow us to disentangle the potential effects of age from that of hearing status. Studies have previously observed that groups of older persons slow down their communication, produce more words, and are more affected by increases in task complexity than younger participants (for review, see Mortensen et al., 2006). Although the studies mentioned in the review by Mortensen and colleagues do not account for potential changes in hearing with age, the findings could suggest that effects of hearing status observed in the current study are driven by the confound between age and hearing status between the groups. If this was the case, we would have expected to find that the older (HI) participants would have significantly slower articulation rates (slower speech), longer IPUs (produce more words), and potentially larger median FTOs. However, in the current study, none of the three measures were affected directly by hearing status (equivalent to age), on the contrary, all measure showed significant, or very near-significant, interactions between hearing status and amplification, indicating that the differences between groups were altered by hearing-aid amplification, which should not be expected if age would be the factor driving the group differences. In a recent study investigating communication between NH and HI interlocutors, no significant effect of age within the HI group was found on the speech levels or the formant changes in speech produced by the HI or their NH interlocutor (age range 53-85 years; Beechey et al., 2020b). For the current study, an exploratory analysis on the effect of age on the outcome measures was done (not shown in Results), showing that within the HI group, age did not significantly predict articulation rate, speech level, or turn-taking timing. It should be noted that the age range for the HI participants in the current study is only 12 years.

5.5 Conclusion

In summary, the current study explored the conversational dynamics between young NH and older HI interlocutors and found that noise increased communication difficulty, especially for participants suffering from hearing impairment. Providing the HI interlocutor with amplification through a hearing aid caused them to behave more like their NH conversational partners, especially in quiet where amplification ensured audibility. These results indicate that hearing-impaired listeners benefit from hearing-aid amplification in quiet situations, even if they do not often report listening difficulties in these situations. The method seems promising for evaluating the benefits of hearing-aid signal processing in more realistic and ecologically valid tests.

5.6 Author contributions

EBP, ENM, and AJMS all contributed to the scoping and designing of the study. EBP conducted the data collection. AJMS and EBP performed the data processing and statistical analysis. EBP, ENM, and AJMS did the interpretation of the results. AJMS and EBP prepared the manuscript.

5.7 Declaration of conflicting interests

The authors declared the following potential conflicts of interest concerning the research, authorship, and/or publication of this article: Eline Borch Petersen is employed at WS Audiology, Lynge, Denmark. No conflicts were declared for the remaining authors.

6 General discussion

This thesis set out to investigate whether objective measures of temporal dynamics and speech production (jointly referred to as *conversational dynamics*) in dialogue could be used to assess conversational difficulty for individuals. This was investigated by establishing baseline behavior between NH interlocutors while conversing in more or less challenging conditions (Chapter 2) and compare this to the pattern of results for NH/HI conversations in increasingly challenging conditions (Chapter 4). We investigated whether simple hearing aid (HA) amplification could alleviate some of the conversational difficulty experienced by HI individuals as indicated by changes in conversational dynamics (Chapter 5). In addition, we investigated the impacts of conversational task (Chapter 3) and face-to-face conversations (Chapter 5) on conversational dynamics.

6.1 Summary of main results

In our first study (Chapter 2), we investigated how noise and L2 impacted conversational dynamics between pairs of NH interlocutors while they solved the DiapixUK task (Baker and Hazan, 2011). In this study and the following two (Chapters 3-4), interlocutors were seated separately, communicating over headphones and microphones. Compared to in quiet, talkers in noise took longer solving the Diapix task, spoke louder, produced longer IPUs, took fewer turns, and answered later and with more variability. Compared to in L1, when conversing in L2, it took participants longer to complete the Diapix task, they spoke slower, produced longer IPUs, took fewer turns, answered faster, and with more variability. We speculated that the longer IPUs, the fewer turns, and the slower articulation functioned as a means to reduce the expected difficulty experienced by the participants when communicating in noise and L2.

In our second study (Chapter 3), in a similar setup to our first study, we investigated the impacts of noise and conversational task on conversational dynamics between pairs of NH interlocutors communicating in their L2. As found in the first study (Chapter 2), in noise, interlocutors produced longer IPUs, answered with more variability, and spoke louder, but they did not answer faster as found in L2 in Chapter 2. Compared to when solving the Diapix task, in free conversation, interlocutors answered faster, spoke faster, produced longer IPUs, and produced more overlaps-within. However, there were no interactions with noise, indicating that people changed their operating point when conversing freely, but the noise impacted the dialogue from this offset in the same manner as when they solved the Diapix task.

In our third study, in a similar setup to the first two, we investigated how noise presented at different levels impacted conversational dynamics between pairs of NH and HI interlocutors. Just as in our first study (Chapter 2), in noise it took participants longer to solve the Diapix task, they spoke louder, produced longer IPUs, took fewer turns, and answered later, and with more variability, all taken as indications of increased conversational effort. Both NH and HI spoke slower in noise, indicating increased speech processing demands, which we did not find when NH conversed with NH individuals. HI individuals produced longer IPUs and answered with more variability than NH individuals, suggesting they were more challenged than their NH partner. Finally, the HI participants spoke more of the time than their NH interlocutor in the loudest noise background, indicating a tendency to dominate the conversational floor.

Our final study investigated the impacts of noise and HA amplification on conversational dynamics between NH/HI interlocutors. In this study, interlocutors were seated face-to-face and heard the noise through loudspeakers instead of headphones. As in our previous studies, in noise, it took participants longer to solve the Diapix task; they spoke louder, produced longer IPUs, took fewer turns, and answered later and with more variability. When the HI participants were compensated for their hearing loss with simple amplification according to the NAL-NL2 prescription in a pair of HAs, they answered with less variability, spoke faster, and produced shorter utterances. These changes in conversational dynamics suggest that the processing was able to alleviate some of the HI participants' communication effort.

6.1.1 The consistency of measures of conversational dynamics

In this section we will discuss the consistency and usefulness of the conversational dynamics common for the four experiments presented in this thesis:

IPUs, FTOs, speech levels and articulation rates.

Interpausal units

Across all four studies, we have found that interlocutors increased their IPU durations in the conditions we designed to be more challenging: talkers produced longer IPUs in noise compared to quiet, and in L2 compared to L1, and HI participants produced longer IPUs than NH talkers. When the HI participants received simple HA compensation in Chapter 5, they reduced their IPU durations to match their NH interlocutor's. We proposed the following explanations for the longer IPUs. If people are experiencing increased demands processing their interlocutor's speech, they may not have planned their entire utterance before they have to launch it at a socially appropriate time. In this case, they could use filler words or repetitions at the phrase boundaries to initiate their turn while still planning it. They could also introduce filled pauses in their turns (Clark and Fox Tree, 2002) to indicate that they are holding the floor while still planning their utterance. These initiatives would naturally increase their IPU durations.

Future studies could analyze the recordings obtained in this thesis for any turn-holding and turn-yielding cues to investigate whether they were more present in utterances produced in noise than in quiet. Acoustic analyses of pitch, loudness, and speech rate changes could be conducted, speech recognition algorithms could analyze whether people had more repetitions and use of filler words, and more sophisticated speech recognition algorithms could search for completions of grammatical clauses.

In our exploratory analysis in Section 2.3.5 we found indications that the longer the IPUs were before floor transfers, the more well-timed the next IPUs were, manifested in lower medians and IQRs of the FTOs. We interpreted this as the responder having more time for speech planning when their partner produced a longer utterance. Conversely, we found that the shorter the IPUs were after floor transfers, the more well-timed they were. This was interpreted as an indication that the planning of a shorter utterance is less demanding than planning a longer one. This is consistent with the results in Chapter 4, where NH interlocutors produced shorter IPUs in the second and third replicates, and also produced shorter and less variable FTOs in these replicates. One could speculate whether these shorter IPUs from NH interlocutors would lead to longer response times for HI interlocutors, given the relationship between IPU durations and

FTOs. If, for example, the NH interlocutors tend to take a more passive role responding more often with one-syllable answers, as was found in Chapter 4, that puts more responsibility on the HI person to find the next difference in the Diapix pair. On the other hand, it decreases the listening demands for the HI participants, consistent with HI individuals' strategy to dominate conversations to reduce the need for listening (Jaworski and Stephens, 1998; Stephens and Zhao, 1996). This is return may put less load on them, which could free up resources for their own planning and thus minimize their FTOs. Since IPU durations increased in all conditions designed to be more challenging, and decreased when HI interlocutors received HA amplification, we believe IPU duration changes are good predictors of difficulty.

Median of floor-transfer offsets

As a measure of the centrality of the FTO distributions, we have used the median. When conversing in their first language, all interlocutors increased their median FTO durations in the presence of noise (Figure 2.7 left panel, Figure 4.7 upper left panel, and Figure 5.5, respectively). In the audio-only experiment (Chapter 4), HI interlocutors had borderline larger median FTO durations than their NH partners in replicates 2 and 3, but not in the first replicate. There was no difference between the median FTOs of the NH and HI participants in the face-to-face experiment (Chapter 5). When talkers conversed in their L2, they decreased their median FTO in noise in the first experiment (Chapter 3), opposite to the pattern seen in L1, but did not change their median FTO in noise in the second experiment (Chapter 4). Thus, the trajectories of the median FTOs, when going from a simpler (quiet, L1) to a more difficult condition (noise, L2) or between NH and HI participants, were not consistent.

We also found that interlocutors' median FTOs decreased in the second experiment (Chapter 3) when conversing freely rather than solving the Diapix task. It was speculated that people could be optimizing for information transfer when solving a task compared to when conversing freely, thus minimizing overlaps (Section 3.4.2), which would result in larger median FTOs in Diapix-elicited dialogue than in free conversation. It was also speculated that the visual search required when solving the Diapix task could delay the response time. As there was no effect of noise on the median FTOs in this experiment, and as the IQRs of FTOs were similar across the two dialog-elicitation methods, we interpreted the effect of task as an offset, a change in operating point, rather than an effect of increased effort in solving the Diapix task compared to conversing freely.

The changes between quiet and noise observed in the NH/NH experiments were minimal compared to the ones observed in NH/HI experiments. In NH/NH conversations in Chapter 2, people answered on average 21 ms later in noise and 19 ms earlier in L2, i.e., very close to 0 ms. In NH/NH conversations in Chapter 3 interlocutors did not change their median FTOs in noise compared to quiet. In this experiment, they also spoke at higher average SNRs than in Chapter 2 (-0.3 dB vs. -2.5 dB SNR), which may have helped them time their turns better, given the negative relationship we found between the SNR their interlocutors communicated at and the participants' ability to time their turns (Section 4.3.5). In NH/HI conversations, both NH and HI interlocutors answered later. In audio-only conversations in Chapter 4 they answered, on average, 115 ms later in 70 dBA SPL noise, and in face-to-face conversations in Chapter 5 they answered, on average, 69 ms later in 70 dB SPL noise. We interpret the audio-only conversations between NH and HI participants (Chapter 4) to be the most challenging for participants, as one of the interlocutors was hearing impaired, and they had no access to visual information. In that experiment, we saw the biggest differences between quiet and noise compared to the other experiments. Together with the fact that participants' delay in turn-taking increased with increasing noise level (Figure 4.7 upper left panel), and that when aided in Chapter 5, HI interlocutors tended to answer faster, this may suggest that the median of FTO durations may be a valuable predictor of communication difficulty, likely in combination with other measures. However, more research is needed to establish the consistency of this measure when people experience increased conversational effort.

One potential reason for why we do not see big differences between NH and HI participants' median FTOs (there was only a difference in the second and third replicate in audio-only conversations, Chapter 4, but no difference in face-to-face conversations, Chapter 5) may be that interlocutors synchronize to target the same average response time (Ten Bosch et al., 2005). Achieving this goal may be more effortful for the HI interlocutor, who we observed to have more variability (FTO IQR) in their timing of turns in both experiments.

IQR of floor-transfer offsets

As a measure of the spread of the FTO distributions, we have used the IQR. Across all four studies, we found that the IQR of FTO distributions increased in the presence of noise. In Chapter 2, the IQR of FTOs increased in L2, and both in Chapter 4 and Chapter 5, HI interlocutors exhibited greater variability in turn-timing than their NH partners. In face-to-face conversations, the offset between NH and HI interlocutors was even larger in noise. However, when HI interlocutors received HA amplification, their IQRs decreased both in quiet and noise. Further, we found a negative relationship between the SNR their partner was communicating at and people's variability of turn-timing.

The pattern of these results all point in the same direction: when it should be more difficult for people to communicate, their timing of turns becomes more variable. As discussed in the general introduction as well as the introductions and discussions in Chapters 2-5, this could be due to people experiencing decreased sensitivity to turn-end prediction cues, making them less precise in their prediction of their interlocutor's turn-end, and be due to interlocutors' increased processing demands due to degraded speech signals. However, follow-up studies are needed to test these hypotheses directly. Objectively analyzing the saliency of turn-end prediction cues would require building models that could analyze the saliency of turn-end prediction cues in the presence of noise or when listening with a hearing loss. Alternatively, studies similar to, e.g., Magyari and Ruiter (2012) and Corps et al. (2018a) could be conducted where people would be listening to turns from the recordings in this thesis and either press a button or verbally respond when they anticipated a turn-end. The drawback of this method is that people are asked to actively perform a task that is automatic in conversation. To analyze increased processing demands, one could use eyetrackers during conversations to analyze pupil dilations of interlocutors as an indicator of effort (Wendt et al., 2017), see Section 6.2.

Speech levels

In all four experiments, interlocutors raised their voice levels substantially when conversing in the presence of noise compared to quiet. In conversations between NH interlocutors, the participants communicated, on average, at negative SNRs (Figures 2.3 and 3.1 right panel). In conversations between NH/HI pairs (Figures 4.4 left panel, and 5.8 left panel), the NH participants communicated, on average, at positive SNRs. The HI, however, communicated at positive SNRs in the audio-only experiment (Chapter 4), but at negative SNRs in the face-to-face experiment (Chapter 5).

NH individuals produced more favorable SNRs when communicating with HI individuals than with NH individuals, likely to increase the audibility for their HI partner. As discussed in Section 5.4.5, NH/HI pairs were seated 2 m apart in the face-to-face experiment, which is a doubling of the distance compared to the calibrated presentation level of their interlocutor in the audio-only NH/HI experiment. Achieving the same SNR at their interlocutor's ear, thus, would require roughly a doubling of their speech levels (Weisser and Buchholz, 2019). As pointed out in Section 4.4.5, people trade off physical effort and comprehension for their interlocutor (Hazan and Baker, 2011). HI interlocutors likely did not have to strain their vocal effort to that extent for their NH interlocutor to understand them in the face-to-face experiment because the NH individuals presumably were able to understand them at negative SNRs, given that NH interlocutors communicated at negative SNRs in the NH/NH conversations. The question is why the HI participants spoke at positive SNRs in the audio-only NH/HI conversations then. People may not only optimize their speech level for their interlocutor but may also adjust it based on their own-voice perception (Laugesen et al., 2009). This may explain the difference between the HI participants' speech levels in the audio-only and face-to-face conversations. Since we estimated the SNR at the interlocutor's position, we do not directly measure the participants' speech levels at their own ears. However, given the differences in (simulated and actual) distance between interlocutors in the audio-only and face-to-face conversations between NH/HI pairs, we expect the speech levels of the interlocutors to be about 6 dB SPL louder at their ears in face-to-face conversations. Thus, even though they produced SNRs that were -0.7 dB on average at their NH interlocutor's ears, they heard their own voice at positive SNRs.

We hypothesized that interlocutors would increase their speech levels in L2 in noise based on studies finding higher SRTs in L2 than L1 (e.g., Wijngaarden et al., 2002), but we found no difference between speech levels in L1 and L2. Wijngaarden et al. (2002) was a listening study, and it is possible that in conversation, people are benefiting from other signal-enhancements such as decreased speech rates (as we found for talkers conversing in L2), or hyperarticulation (which we did not measure in this thesis) than just SNR adjustments. There is a higher degree of context in real conversation than in listening tasks, and in our experiment, interlocutors were given a language-independent context from the Diapix pairs. Further, the interlocutors had matched accents as they shared the same L1, which could facilitate prediction of the content of their speech. In Section 2.4.4 we speculated further on the lack of differences in SNR between languages.

Thus, the speech levels depend on the means of communication (face-to-face or audio-only) and the interlocutor. We found in Chapter 4 that there was a negative correlation between participants' median and IQR of FTO durations and the SNR their partner communicated at (see Figures 4.8 and A.3). Thus, this indicates that the level people communicate at impacts their interlocutor's ability to time their turns. This could indicate that SNR can be used as a predictor of the other person's difficulty.

Articulation rates

In the NH/NH conversations, interlocutors' articulation rates were unaffected by background noise. When speaking in L2, talkers spoke slower, consistent with articulation rate being an indicator of fluency in language (De Jong and Wempe, 2009; García Lecumberri et al., 2017). In audio-only NH/HI conversations in Chapter 4, both NH and HI talkers spoke slower in the presence of noise and even slower in the highest noise level compared to lower noise levels. In quiet and the lowest noise level, NH talkers spoke quicker than their HI partners, whereas in the loudest noise levels, there was no difference between NH and HI participants' articulation rates. This was interpreted as the NH interlocutors adapting their rates to alleviate the increased speech processing difficulties for the HI interlocutors in noise.

In face-to-face conversations in Chapter 5, there was no difference between NH and HI talkers, and both groups had lower articulation rates in noise in the first replicate. In the second replicate, they spoke significantly slower in quiet, whereas their articulation rates in noise were unchanged, leaving no difference between quiet and noise. However, when the HI participants received HA compensation, they spoke significantly faster.

The results are not entirely consistent across all our studies. Levitan and Hirschberg (2011) showed that interlocutors adapted their speech rates to their partners by showing that their proximity to their partner's speaking rate was higher than to others with whom they did not converse. They also showed that people's speaking rate was more similar to their own in another session than to their partner's, suggesting some personal speaking behavior. Hazan and Baker (2011) found that NH talkers adapted their speech rates to the difficulty their partner experienced. NH interlocutors in audio-only NH/HI conversations may have exhibited more personal speaking behavior in the conditions without background noise and with low background noise, explaining the offset between NH and HI interlocutors in those conditions, but had to adapt to their interlocutor's difficulty parsing their speech in louder background noises. In face-to-face communication, they may have synchronized more with their HI partner. As the age spans of the NH and the HI groups across the NH/HI studies were the same, age should not be able to explain the observed differences between those studies. It is unclear where the differences come from, and further studies are needed to establish the effects of conversation difficulty on articulation rates. However, the fact that NH talkers spoke slower in L2, that NH talkers tended to speak slower when talking to an HI partner rather than an NH partner, that interlocutors spoke slower in higher noise levels, and that HI interlocutors spoke faster when receiving compensation for their hearing loss indicates that articulation rate is a valuable predictor of difficulty. More research, however, is needed to establish the consistency of the measure.

6.1.2 Are NH individuals experiencing more effort when conversing with an HI interlocutor?

The NH participants on average increased their median FTOs by, on average, 129 ms and their IQR of FTOs by, on average, 182 ms when conversing in the presence of 70 dBA SPL noise compared to quiet in audio-only NH/HI conversations. In face-to-face NH/HI conversations their median and IQR of FTOs increased, on average, by 64 and 98 ms, respectively. In comparison, their medians and IQRs of FTOs only increased by 21 ms and 41 ms, respectively, in NH/NH conversations (Chapter 2). Whereas there was no difference between the median of the NH and HI interlocutors' FTOs, the HI had significantly bigger variability (IQR) in their turn-timing. Anecdotally, in informal interviews with participants after the experiment, NH participants seemed more tired after conducting the experiment with an HI partner (Chapter 4) than when conversing with an NH partner (Chapter 2). As discussed in the introduction, conversing is a dynamic feedback process between interlocutors. To facilitate smooth conversation, people synchronize their behavior and use signal-enhancement strategies to increase the predictability of their speech (Donnarumma et al., 2017; Garrod and Pickering, 2004). Pouw and Holler (2020) found lag-1 autocorrelations in FTO series suggesting that people monitor the other person's FTO and adjust their own FTO accordingly a turn later. Thus, if the NH interlocutors monitor their HI partner's FTO which is more variable, as we found in both NH/HI experiments, the NH individual may experience increased effort trying to adapt to their FTOs. The NH talkers in the NH/HI experiments spoke slower and louder to their HI interlocutor than NH talkers did in NH/NH conversations. This can be taken as indications the NH participants had to work harder to increase the saliency of their speech to compensate for their interlocutor's decreased sensitivity to cues that would normally be enough to reach coherence between conversational partners.

Levitan and Hirschberg (2011) presents a framework for measuring speech feature entrainment between interlocutors by measuring proximity, convergence and synchrony between interlocutors on various dimensions of speech production. This framework and the lag-1 autocorrelation methodology presented in Pouw and Holler (2020) could be applied to the datasets obtained in this thesis to study the role of adaptation and synchronization between interlocutors.

6.1.3 Task vs. free conversation

We used the Diapix task to elicit dialogue between interlocutors in our experiments. The advantage of using this task is that it is possible to measure the task completion time as a proxy of communication efficiency (Baker and Hazan, 2011; Van Engen et al., 2010). Further, it provides more control over the experiment, as the context of the conversation is the same for all conversational pairs. Even though interlocutors in the free conversations in Chapter 3 were given conversational topics, it can be awkward to converse with an unacquainted interlocutor, which could introduce variability in their conversational dynamics. The question is whether the conversational dynamics observed in Diapix-elicited dialogues are representative of dynamics in free conversation. While we saw offset differences between some of the conversational dynamics measures between the dialogue-elicitation methods in Chapter 3, we saw no interactions between noise and the dialogue-elicitation method on any of the measures. This suggests that it is possible to generalize the impacts of noise in NH/NH conversations from Diapix-elicited dialog compared to more free conversation. However, we did not inspect the impacts of task on dialogue involving HI interlocutors. HI interlocutors may rely on the context given from solving the task to a greater extent than NH interlocutors, and noise may therefore affect them more in free conversation where the context is not given from the task. The differences may also be greater between Diapix-elicited dialogue and free conversation in face-to-face conversations, as the Diapix task forces people to look down on their Diapix, which may make them less able to make use of their partner's gestures and facial expressions to enhance speech understanding. When people are separated in booths, they cannot utilize these cues regardless of whether they are solving a task or not. Further research is needed to establish the impacts of task on conversational dynamics in dialogue involving HI participants and in face-to-face conversations.

6.1.4 Age vs. hearing loss

The NH/HI studies presented in this thesis were not designed to disentangle the effects of hearing loss and age. The average age of participants in the NH groups was much lower than that of the HI group. We chose to recruit these groups for several reasons. First, we wanted to establish baseline behavior between young, cognitively healthy participants with normal hearing without any cognitive decline and thus chose university students or recently graduated students in their mid-20s for our first two experiments. For our next two studies, we wanted to compare how people recruited with the same criteria would behave when conversing with HI interlocutors. Even though the experiments were not designed to disentangle the effects of noise and hearing loss, we saw some indications that the effects we observed were driven by hearing loss and not age. In Chapter 4, age was not a significant predictor of any of the conversational dynamics. As we pointed out in Section 4.4.6, this null-result should not be taken as evidence against age as an effect. In Chapter 5, HA amplification affected almost all conversational dynamics measures in the direction opposite to that observed in noise, which was taken as an indication that amplification alleviated the HI participants from some of their increased conversational load due to their hearing loss.

A fifth study was designed to be able to disentangle the effects of hearing loss and age. Unfortunately, due to the COVID-19 lockdown in Denmark, we were not able to conduct this experiment. The study was designed to investigate the effects of noise on conversational dynamics between triplets consisting of pairs of 1) young NH (YNH) interlocutors, 2) YNH and older NH (ONH) interlocutors, and 3) ONH and older HI (OHI) interlocutors. By comparing how YNH interlocutors adapted to their ONH and OHI partners, respectively, it would be possible to attribute the adaptions to either hearing loss or age of their interlocutor. Similarly, by comparing ONH and OHI's behavior, it would be possible to attribute the changes observed as a consequence of noise interference to either hearing loss or age.

6.1.5 Face-to-face vs. audio-only conversations

The changes between quiet and noise were more enhanced in the audio-only conversations (Chapters 4) than in face-to-face conversations between NH/HI pairs (Chapter 5), likely because people benefited from the added visual modality. The advantage of studying people in audio-only setups is that their communication channel is restricted to audio only, making the experiment more controlled. This prevents them from using the visual modality to enhance their speech signal, by fx using gestures, and it prevents them from being able to do lip-reading. Thus, it is easier to measure the effect of their hearing in isolation without having variability introduced by how good they are at utilizing non-acoustic compensatory strategies. While the effects of noise on articulation rate differed between audio-only and face-to-face conversations between NH/NH and NH/HI pairs, the rest of the conversational dynamics measured showed the same direction in the presence of noise compared to quiet, suggesting it is possible to generalize audio-only interactions to face-to-face interactions to some extent. Some of the variability in face-to-face conversations compared to audio-only conversations may be accounted for by capturing other behavior than those emerging acoustically (see Section 6.2).

6.2 The future of conversational dynamics in hearing research

An exciting prospect of the proposed method for measuring interaction is that it is objective and can be measured online as conversations unfold, making it possible to incorporate these measures on a hearing assistive device (HAD). Participants are asked to perform an automatic, natural task, i.e., conversing, so they do not need instructions or training to log relevant data on an HAD. Logging such metrics on-the-fly would make it possible to do real-time adjustments to the needs of the wearer. However, a first step towards this is to establish what changes in conversational dynamics signal difficulty. As mentioned in Section 6.1.5, the changes between quiet and noise were more enhanced in audio-only conversations between NH/HI interlocutors, likely because people were able to utilize the visual modality for signal enhancement in face-to-face conversations and thereby building coherence. Thus, it would be relevant to capture these enhancements along with other biomarkers indicating difficulty in conversation.

At Facebook Reality Labs (FRL), we built a lab that was able to capture multimodal behavior data in a spatialized array of 52 loudspeakers (Sørensen and Brimijoin, 2020). We recorded people's voices as in the experiments presented in this thesis. Beyond that, we captured people's pupil dilations and gaze patterns on eye trackers and captured head and torso movements with a motion capture system. In a similar setup, Hadley et al. (2019) found that people gazed more towards their interlocutor's mouth in noise conditions, suggesting they were utilizing lip-reading. Capturing pupil dilations can provide insights into the effort people are spending listening (e.g., Wendt et al., 2017), and together with measures of skin conductance as a stress indicator and delays and variability in turn-taking, these measures could indicate the difficulty the person is experiencing communicating. From the data captured with this setup, we used turn-taking points to analyze behavior in all sensors captured leading up to, during, and following turn-taking. Not a lot of research has focused on capturing multimodal behavior, and more research is needed to establish the consistency of the dynamic measures when people are experiencing difficulty in conversation to be able to use it as a tool for diagnostics and on-the-fly adjustments to HADs to alleviate people's difficulty participating in conversation. This thesis took a first step towards that.

The dynamic measures outlined above could also be used to decode whom a person is trying to converse with. In preliminary data between triads captured with our setup at FRL, we found that when a person talked, their interlocutors typically gazed directly at them, whereas the talker shifted his/her gaze between his/her interlocutors. In exploratory analyses conducted in this thesis (but not presented), we found that there were substantial differences between conversational dynamics produced between interlocutors, and dynamics between two random talkers, as was also shown by Aubanel et al. (2012) and Levitan and Hirschberg (2011). This information could be used to steer a beamformer on a person's HAD towards the person the HAD detected to be the wearer's target conversational partner to enhance the speech from this partner.

The results of the studies presented in this thesis also has the potential to be used beyond the fields of hearing research. Fx, the knowledge about how people communicate and how that changes in the presence of noise could be used in the development of artificial conversational partners (Heldner and Edlund, 2010).

A

Appendix

A.1 Supplementary material

A.1.1 Floor-transfer offset vs. interpausal unit duration, Experiment 1

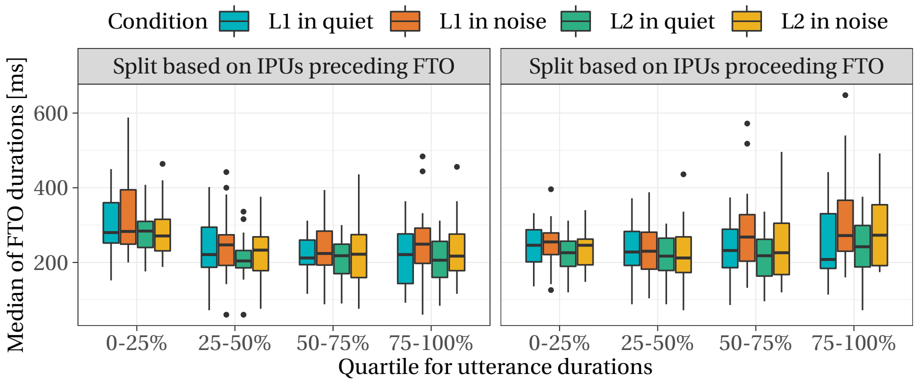

Figure A.1: Boxplots of median of floor-transfer offsets (FTOs) divided by four percentiles of the preceding utterance duration (left panel) and following utterance duration (right panel) for the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2.

Predictive variable	Contrast	L1 in quiet	L1 in noise	L2 in quiet	L2 in noise
Median of IPU duration before floor-transfer	Q1-Q2	[$t(57) = 4.8$, p <.001] ***	[$t(57) = 6$, p <.001] ***	[$t(57) = 7.2$, p <.001] ***	[$t(57) = 3.7$, p <.001] ***
	Q1-Q3	[$t(57) = 5.6$, p <.001] ***	[$t(57) = 6.8$, p <.001] ***	[$t(57) = 7.1$, p <.001] ***	[$t(57) = 4.2$, p <.001] ***
	Q1-Q4	[$t(57) = 5.9$, p <.001] ***	[$t(57) = 5.2$, p <.001] ***	[$t(57) = 7.8$, p <.001] ***	[$t(57) = 3.6$, p <.001] ***
	Q2-Q3	[$t(57) = .87$, $p = .39$]	[$t(57) = .81$, $p = .42$]	[$t(57) = -.08$, $p = .93$]	[$t(57) = .48$, $p = .63$]
	Q2-Q4	[$t(57) = 1.1$, $p = .26$]	[$t(57) = -.8$, $p = .43$]	[$t(57) = .57$, $p = .57$]	[$t(57) = -.13$, $p = .9$]
	Q3-Q4	[$t(57) = .26$, $p = .78$]	[$t(57) = -1.6$, $p = .11$]	[$t(57) = .66$, $p = .51$]	[$t(57) = -.61$ $p = .54$]
Median of IPU duration after floor-transfer	Q1-Q2	[$t(57) = .66$, $p = .51$]	[$t(57) = 1.4$, $p = .18$]	[$t(57) = .87$, $p = .39$]	[$t(57) = 1.3$, $p = .2$]
	Q1-Q3	[$t(57) = .5$, $p = .62$]	[$t(57) = -2.1$, $p <.05$] *	[$t(57) = 1.4$, $p = .15$]	[$t(57) = -.52$, $p = .6$]
	Q1-Q4	[$t(57) = -.87$, $p = .39$]	[$t(57) = -4.6$, $p <.01$] **	[$t(57) = -1.5$, $p = .14$]	[$t(57) = -3.8$, $p <.001$] ***
	Q2-Q3	[$t(57) = -.17$, $p = .87$]	[$t(57) = -3.4$, $p <.01$] **	[$t(57) = .58$, $p = .56$]	[$t(57) = -1.8$, $p = .076$] .
	Q2-Q4	[$t(57) = -1.5$, $p = .13$]	[$t(57) = -5.9$, $p <.001$] ***	[$t(57) = -2.4$, $p <.05$] *	[$t(57) = -5$, $p <.001$] ***
	Q3-Q4	[$t(57) = -1.4$, $p = .18$]	[$t(57) = -2.5$, $p <.01$] *	[$t(57) = -2.9$, $p <.01$] **	[$t(57) = -3.2$, $p <.01$] **
IQR of IPU duration before floor-transfer	Q1-Q2	[$t(57) = 1.3$, $p = .19$]	[$t(57) = 3.8$, $p <.001$] ***	[$t(57) = -.19$, $p = .85$]	[$t(57) = 1.7$, $p = .09$] .
	Q1-Q3	[$t(57) = 5$, $p <.001$] ***	[$t(57) = 5.7$, $p <.001$] ***	[$t(57) = 2.2$, $p <.05$] *	[$t(57) = 3.7$, $p <.001$] ***
	Q1-Q4	[$t(57) = 4.3$, $p <.001$] ***	[$t(57) = 5.3$, $p <.001$] ***	[$t(57) = 3.8$, $p <.001$] ***	$t(57) = 2.2$, $p <.01$] *
	Q2-Q3	[$t(57) = 3.7$, $p <.001$] ***	[$t(57) = 1.9$, $p = .06$] .	[$t(57) = 2.4$, $p <.05$] *	[$t(57) = 1.9$, $p = .06$] .
	Q2-Q4	[$t(57) = 3$, $p <.01$] **	[$t(57) = 1.5$, $p = .15$]	[$t(57) = 4$, $p <.001$] ***	[$t(57) = .45$, $p = .65$]
	Q3-Q4	[$t(57) = -.75$, $p = .46$]	[$t(57) = -.44$, $p = .66$]	[$t(57) = 1.5$, $p = .13$]	[$t(57) = -1.5$, $p = .14$]
IQR of IPU duration after floor-transfer	Q1-Q2	[$t(57) = -5$, $p <.001$] ***	[$t(57) = -4$, $p <.001$] ***	[$t(57) = -5.4$, $p <.001$] ***	[$t(57) = -6.8$, $p <.001$] ***
	Q1-Q3	[$t(57) = -10$, $p <.001$] ***	[$t(57) = -8.2$, $p <.001$] ***	[$t(57) = -10$, $p <.001$] ***	[$t(57) = -8.8$, $p <.001$] ***
	Q1-Q4	[$t(57) = -12$, $p <.001$] ***	[$t(57) = -11$, $p <.001$] ***	[$t(57) = -10$, $p <.001$] ***	[$t(57) = -10.6$, $p <.001$] ***
	Q2-Q3	[$t(57) = -5.3$, $p <.001$] ***	[$t(57) = -4.2$, $p <.001$] ***	[$t(57) = -5$, $p <.001$] ***	[$t(57) = -2$, $p <.05$] *
	Q2-Q4	[$t(57) = -6.8$, $p <.001$] ***	[$t(57) = -7$, $p <.001$] ***	[$t(57) = -4.8$, $p <.001$] ***	[$t(57) = -3.8$, $p <.001$] ***
	Q3-Q4	[$t(57) = -1.5$, $p = .14$]	[$t(57) = -2.8$, $p <.01$] **	[$t(57) = .23$, $p = .82$]	[$t(57) = -1.8$, $p = .078$] .

Table A.1: Pairwise comparisons of least-squares means between neighboring quartiles (stated in the “Contrast” column). Mixed effects regression models were fitted to the predictive variable in the leftmost column.

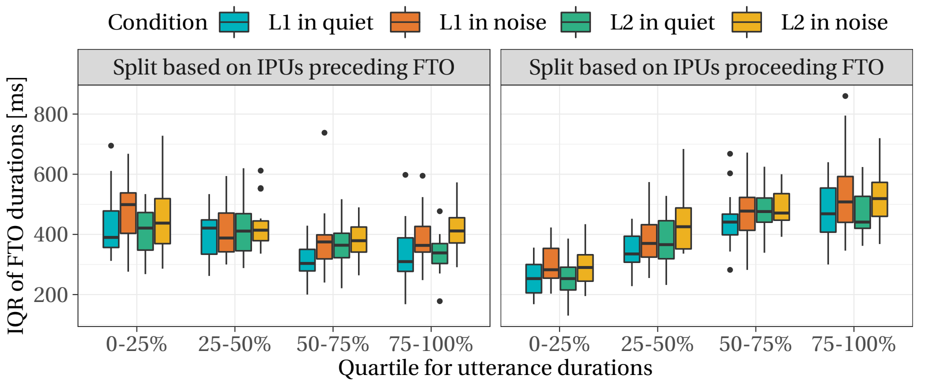

Figure A.2: Boxplots of interquartile-range of floor-transfer offsets (FTOs) divided by four percentiles of the preceding utterance duration (left panel) and following utterance duration (right panel) for the four conditions: quiet in first language (L1), quiet in second language (L2), noise in L1, and noise in L2.

A.1.2 Median floor-transfer offset vs. partner SNR, Experiment 3

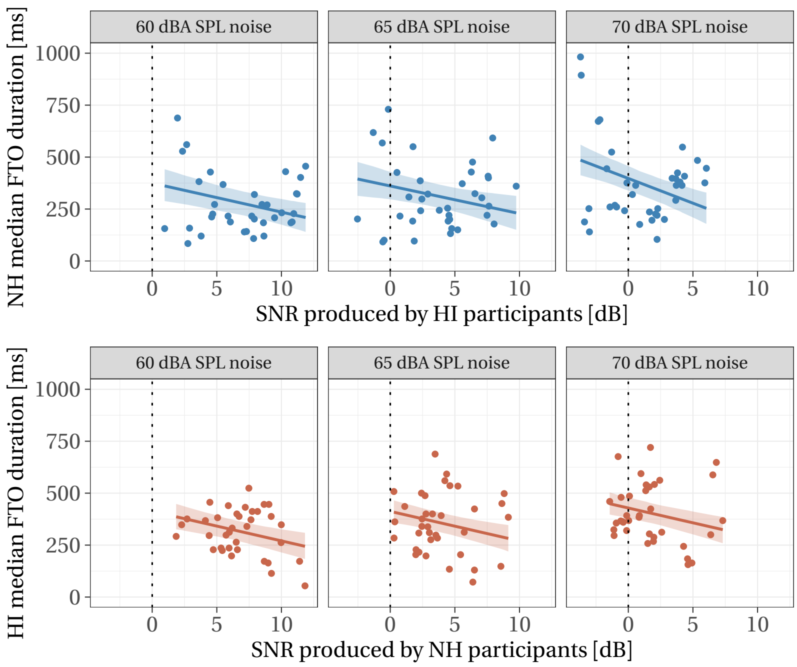

Figure A.3: Raw medians of FTO durations (points) as a function of the SNR received by the interlocutor that took over the floor in the three noise conditions: 60 dBA SPL, 65 dBA SPL, and 70 dBA SPL; data are shown with regression lines and 95% prediction intervals. Data are shown for each participant and replicate for NH participants as a function of the SNR produced by their HI interlocutors (upper panel) and for the HI participants as a function of the SNR produced by their NH interlocutors (lower panel).

Bibliography

Aubanel, V. and M. Cooke (2013). "Strategies adopted by talkers faced with fluctuating and competing-speech maskers". In: *The Journal of the Acoustical Society of America* 134.4, pp. 2884–2894.

Aubanel, V., M. Cooke, E. Foster, M. L. Garcia Lecumberri, and C. Mayo (2012). "Effects of the availability of visual information and presence of competing conversations on speech production". In: *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012* 3.September, pp. 2031–2034.

Aubanel, V., M. Cooke, J. Villegas, and M. L. Garcia Lecumberri (2011). "Conversing in the presence of a competing conversation: Effects on speech production". In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* May 2014, pp. 2833–2836.

Baker, R. and V. Hazan (2011). "DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs". In: *Behavior Research Methods* 43.3, pp. 761–770.

Barthel, M., A. S. Meyer, and S. C. Levinson (2017). "Next speakers plan their turn early and speak after turn-final "go-signals"". In: *Frontiers in Psychology* 8.393.

Barthel, M. and S. Sauppe (2019). "Speech Planning at Turn Transitions in Dialog Is Associated With Increased Processing Load". In: *Cognitive Science* 43.7, pp. 1–16.

Barthel, M., S. Sauppe, S. C. Levinson, and A. S. Meyer (2016). "The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm". In: *Frontiers in Psychology* 7.DEC, pp. 1–13.

Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker (2015). "Fitting linear mixed-effects models using lme4". In: *Journal of Statistical Software* 67.1.

Beechey, T., J. M. Buchholz, and G. Keidser (2018). “Measuring communication difficulty through effortful speech production during conversation”. In: *Speech Communication* 100.April, pp. 18–29.

Beechey, T., J. M. Buchholz, and G. Keidser (2019). “Eliciting naturalistic conversations: A method for assessing communication ability, subjective experience, and the impacts of noise and hearing impairment”. In: *Journal of Speech, Language, and Hearing Research* 62.2, pp. 470–484.

Beechey, T., J. M. Buchholz, and G. Keidser (2020a). “Hearing aid amplification reduces communication effort of people with hearing impairment and their conversation partners”. In: *Journal of Speech, Language, and Hearing Research* 63.4, pp. 1299–1311.

Beechey, T., J. M. Buchholz, and G. Keidser (2020b). “Hearing impairment increases communication effort during conversations in noise”. In: *Journal of Speech, Language, and Hearing Research* 63.i, pp. 1–24.

Beilin and A. Powers (2013). *True Advances in Hearing Aid Technology: What Are They and Where’s the Proof?*

Bekke, M. ter, L. Drijvers, and J. HOLLER (2020). “The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech”. In: *Gesture and Speech in Interaction (GESPIN2020)*, pp. 1–6.

Bisgaard, N., M. S. Vlaming, and M. Dahlquist (2010). “Standard Audiograms for the IEC 60118-15 Measurement Procedure”. In: *Trends in Amplification* 14.2, pp. 113–120.

Boersma, P. and D Weenink (2017). *Praat: Doing Phonetics by Computer.*

Bögels, S., K. H. Kendrick, and S. C. Levinson (2015a). “Never say no . . . How the brain interprets the pregnant pause in conversation”. In: *PLoS ONE* 10.12, pp. 1–15.

Bögels, S., L. Magyari, and S. C. Levinson (2015b). “Neural signatures of response planning occur midway through an incoming question in conversation”. In: *Scientific Reports* 5, pp. 1–11.

Bögels, S. and F. Torreira (2015). “Listeners use intonational phrase boundaries to project turn ends in spoken interaction”. In: *Journal of Phonetics* 52, pp. 46–57.

Boiteau, T. W., P. S. Malone, S. A. Peters, and A. Almor (2014). “Interference between conversation and a concurrent visuomotor task”. In: *J Exp Psychol Gen.* 143.1, pp. 295–311.

Brady, P. T. (1968). “A Statistical Analysis of On-Off Patterns in 16 Conversations”. In: *Bell System Technical Journal* 47.1, pp. 73–91.

Brimijoin, W. O., D. McShefferty, and M. A. Akeroyd (2012). “Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task”. In: *Hearing Research* 283.1-2, pp. 162–168.

Brusco, P., J. Vidal, Š. Beňuš, and A. Gravano (2020). “A cross-linguistic analysis of the temporal dynamics of turn-taking cues using machine learning as a descriptive tool”. In: *Speech Communication* 125.September, pp. 24–40.

Carlile, S. and G. Keidser (2020). “Conversational Interaction Is the Brain in Action: Implications for the Evaluation of Hearing and Hearing Interventions”. In: *Ear and hearing* 41.Ear Hear. 2020 Nov/Dec, 56S–67S.

Casillas, M., S. C. Bobb, and E. V. Clark (2015). “Turn-taking, timing, and planning in early language acquisition”. In: *Journal of Child Language*, pp. 1–28.

Clark, H. H. and J. E. Fox Tree (2002). “Using uh and um in spontaneous speaking”. In: *Cognition* 84.1, pp. 73–111.

Cooke, M., S. King, M. Garnier, and V. Aubanel (2014). “The listening talker: A review of human and algorithmic context-induced modifications of speech”. In: *Computer Speech and Language* 28.2, pp. 543–571.

Corps, R. E., A. Crossley, C. Gambi, and M. J. Pickering (2018a). “Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it”. In: *Cognition* 175.March, pp. 77–95.

Corps, R. E., C. Gambi, and M. J. Pickering (2018b). “Coordinating Utterances During Turn-Taking: The Role of Prediction, Response Preparation, and Articulation”. In: *Discourse Processes* 55.2, pp. 230–240.

De Jong, N. H. and T. Wempe (2009). “Praat script to detect syllable nuclei and measure speech rate automatically.” In: *Behavior research methods* 41.2, pp. 385–90.

De Ruiter, J. P., H. Mitterer, and N. J. Enfield (2006). “Projecting the end of a speaker’s turn: A cognitive cornerstone of conversation”. In: *Language* 82.3, pp. 515–535.

Donnarumma, F., H. Dindo, P. Iodice, and G. Pezzulo (2017). “You cannot speak and listen at the same time: a probabilistic model of turn-taking”. In: *Biological Cybernetics* 111.2, pp. 165–183.

Draper, M. H., P. Ladefoged, and D. Whitteridge (1960). “Expiratory Pressures and Air Flow During Speech”. In: *British Medical Journal* 1.5189, pp. 1837–1843.

Dreschler, W., H. Verschuure, C. Ludvigsen, and S. Westermann (2001). “ICRA Noises: Artificial Noise Signals with Speech-like Spectral and Temporal Properties for Hearing Instrument Assessment: Ruidos ICRA: Señates de ruido artificial con espectro similar al habla y propiedades temporales para pruebas de instrumentos auditiv”. In: *International Journal of Audiology* 40.3, pp. 148–157.

Duncan, S. et al. (1972). “Some Signals and Rules for Taking Speaking”. In: *Journal of Personality and Social Psychology* 23.2, pp. 283–292.

Fitzpatrick, M., J. Kim, and C. Davis (2015). “The effect of seeing the interlocutor on auditory and visual speech production in noise”. In: *Speech Communication* 74, pp. 37–51.

García Lecumberri, M. L., M. Cooke, and M. Wester (2017). “A bi-directional task-based corpus of learners’ conversational speech”. In: *International Journal of Learner Corpus Research* 3.2, pp. 175–195.

Garnier, M., N. Henrich, and D. Dubois (2010). “Influence of sound immersion and communicative interaction on the lombard effect”. In: *Journal of Speech, Language, and Hearing Research* 53.3, pp. 588–608.

Garrod, S. and M. J. Pickering (2004). “Why is conversation so easy?” In: *Trends in Cognitive Sciences* 8.1, pp. 8–11.

Giles, H., N. Coupland, and J. Coupland (1991). *Accommodation theory: Communication, context, and consequence.* Cambridge University Press, pp. 1–68.

Gisladottir, R., S. Levinson, and D. Chwilla (2015). “EEG data for "Conversation electrified: ERP correlates of speech act recognition in underspecified utterances"”. In: *Plos One*, pp. 1–24.

Golestani, N., S. Rosen, S. K. Scott, S. K. Scoot, S. K. Scott, and S. K. Scoot (2009). “Native-language benefit for understanding speech-in-noise: The contribution of semantics”. In: *Language and Cognition* 12.3, pp. 385–392.

Grant, K. W. (1987a). “Encoding voice pitch for profoundly hearing-impaired listenersa)”. In: *Journal of the Acoustical Society of America* 82.2, pp. 423–432.

Grant, K. W. (1987b). “Identification of intonation contours by normally hearing and profoundly hearing-impaired listenersa)”. In: *Journal of the Acoustical Society of America* 82.4, pp. 1172–1178.

Gravano, A. (2010). “Turn-taking and affirmative cue words in task-oriented dialogue”. In: *Dissertation Abstracts International, B: Sciences and Engineering* 70.8, p. 4943.

Gravano, A. and J. Hirschberg (2011). “Turn-taking cues in task-oriented dialogue”. In: *Computer Speech and Language* 25.3, pp. 601–634.

Hadley, L. V., W. O. Brimijoin, and W. M. Whitmer (2019). “Speech, movement, and gaze behaviours during dyadic conversation in noise”. In: *Scientific Reports* 9.1, pp. 1–8.

Hadley, L. V., N. K. Fisher, and M. J. Pickering (2020a). “Listeners are better at predicting speakers similar to themselves”. In: *Acta Psychologica* 208.June, p. 103094.

Hadley, L. V., W. M. Whitmer, W. O. Brimijoin, and G. Naylor (2020b). “Conversation in small groups: Speaking and listening strategies depend on the complexities of the environment and group”. In: *Psychonomic Bulletin and Review*.

Hagoort, P. and P. Indefrey (2014). “The neurobiology of language beyond single words”. In: *Annual Review of Neuroscience* 37, pp. 347–362.

Hazan, V. and R. Baker (2011). “Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions”. In: *The Journal of the Acoustical Society of America* 130.4, pp. 2139–2152.

Hazan, V., O. Tuomainen, J. Kim, C. Davis, P. Sciences, and W. Sydney (2019). “The effect of visual cues on speech characteristics of older and younger adults in an interactive task”. In: *ICPhS 2019*, pp. 815–819.

Hazan, V., O. Tuomainen, J. Kim, C. Davis, B. Sheffield, and D. Brungart (2018a). “Clear speech adaptations in spontaneous speech produced by young and older adults”. In: *The Journal of the Acoustical Society of America* 144.3, pp. 1331–1346.

Hazan, V. et al. (2018b). “How do aging and age-related hearing loss affect the ability to communicate effectively in challenging communicative conditions?” In: *Hearing Research* 369, pp. 33–41.

Heldner, M. and J. Edlund (2010). “Pauses, gaps and overlaps in conversations”. In: *Journal of Phonetics* 38.4, pp. 555–568.

Hengen, J., I. L. Hammarström, and S. Stenfelt (2020). “Perception of One’s Own Voice After Hearing-Aid Fitting for Naive Hearing-Aid Users and Hearing-Aid Refitting for Experienced Hearing-Aid Users”. In: *Trends in Hearing* 24.

Hjalmarsson, A. (2011). “The additive effect of turn-taking cues in human and synthetic voice”. In: *Speech Communication* 53.1, pp. 23–35.

Hornsby, B. W., G. Naylor, and F. H. Bess (2016). "A taxonomy of fatigue concepts and their relation to hearing loss". In: *Ear and Hearing*. Vol. 37. Suppl 1, 136S–144S.

Høydal Harry, E. (2017). "A new own voice processing system for optimizing communication". In: *Hearing Review* 24.11, pp. 20–22.

Indefrey, P. and W. J. Levelt (2004). "The spatial and temporal signatures of word production components". In: *Cognition* 92.1-2, pp. 101–144.

Jaworski, A. and D. Stephens (1998). "Self-reports on silence as a face-saving strategy by people with hearing impairment". In: *International Journal of Applied Linguistics (United Kingdom)* 8.1, pp. 61–80.

Junqua, J. C. (1996). "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex". In: *Speech Communication* 20.1-2, pp. 13–22.

Kalathottukaren, R. T., R. Purdy, S. C. McCormick, and E. Ballard (2015). "Behavioral Measures to Evaluate Prosodic Skills: A Review of Assessment Tools for Children and Adults". In: *Contemporary Issues in Communication Science and Disorders* 42.Spring, pp. 138–154.

Keidser, G., H. Dillon, L. Carter, and A. O'Brien (2012). "NAL-NL2 Empirical Adjustments". In: *Trends in Amplification* 16.4, pp. 211–223.

Keidser, G. et al. (2020). "The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it". In: *Ear and Hearing* 41, 5S–19S.

Kendrick, K. H. and F. Torreira (2015). "The Timing and Construction of Preference: A Quantitative Study". In: *Discourse Processes* 52.4, pp. 255–289.

Kiessling, J. et al. (2003). "Candidature for and delivery of audiological services: Special needs of older people". In: *International Journal of Audiology* 42.SUPPL. 2.

Kuk, F., C. Slugocki, N. Ruperto, and P. Korhonen (2021). "Performance of normal-hearing listeners on the Repeat-Recall test in different noise configurations". In: *International Journal of Audiology* 60.1, pp. 35–43.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen (2017). "lmerTest Package: Tests in Linear Mixed Effects Models". In: *Journal of Statistical Software* 82.13, pp. 1–26.

Laugesen, S., C. Nielsen, P. Maas, and N. S. Jensen (2009). "Observations on hearing aid users' strategies for controlling the level of their own voice". In: *Journal of the American Academy of Audiology* 20.8, pp. 503–513.

Levinson, S. C. and F. Torreira (2015). “Timing in turn-taking and its implications for processing models of language”. In: *Frontiers in Psychology* 6.JUN, pp. 1–17.

Levitan, R. and J. Hirschberg (2011). “Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions”. In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* August, pp. 3081–3084.

Li, X. and F.-C. Jeng (2011). “Noise tolerance in human frequency-following responses to voice pitch”. In: *The Journal of the Acoustical Society of America* 129.1, EL21–EL26.

Lind, C. (2012). “Living Well: Everyday Conversation and Hearing Impairment”. In: *Hearing Review* 19, 16–18 3p.

Lunner, T., M. Rudner, T. Rosenbom, J. Ågren, and E. H. N. Ng (2016). “Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions”. In: *Ear and Hearing*. Vol. 37, 145S–154S.

Magyari, L., M. C. M. Bastiaansen, J. P. De Ruiter, and S. C. Levinson (2014). “Early Anticipation Lies behind the Speed of Response in Conversation”. In: *Journal of Cognitive Neuroscience* 26, pp. 2530–2539.

Magyari, L. and J. P. de Ruiter (2012). “Prediction of turn-ends based on anticipation of upcoming words”. In: *Frontiers in Psychology* 3.OCT, pp. 1–9.

Mansour, N. (2021). *Assessing hearing device benefit using virtual sound environments*. Vol. 46. Technical University of Denmark.

Mansour, N., M. Marschall, T. May, A. Westermann, and T. Dau (2021). “A method for realistic, conversational signal-to-noise ratio estimation”. In: *The Journal of the Acoustical Society of America* 149.3, pp. 1559–1566.

Mattys, S. L., J. Brooks, and M. Cooke (2009). “Recognizing speech under a processing load: Dissociating energetic from informational factors”. In: *Cognitive Psychology* 59.3, pp. 203–243.

Mayo, L. H., M. Florentine, and S. Buus (1997). “Age of second-language acquisition and perception of speech in noise”. In: *Journal of Speech, Language, and Hearing Research* 40.3, pp. 686–693.

McGarrigle, R. et al. (2014). *Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'*.

McInerney, M. and P. Walden (2013). “Evaluating the use of an assistive listening device for communication efficiency using the diapix task: A pilot study”. In: *Folia Phoniatrica et Logopaedica* 65.1, pp. 25–31.

Menenti, L., S. M. Gierhan, K. Segaert, and P. Hagoort (2011). “Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI”. In: *Psychological Science* 22.9, pp. 1173–1182.

Mertens, J. and J. P. de Ruiter (2021). “Cognitive and social delays in the initiation of conversational repair”. In: *Dialogue and Discourse* 12.1, pp. 21–44.

Mirkovic, B., S. Debener, J. Schmidt, M. Jaeger, and T. Neher (2019). “Effects of directional sound processing and listener’s motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ?” In: *Hearing Research* 377, pp. 260–270.

Moore, B. C. (1987). “Psychophysics of normal and impaired hearing”. In: *British Medical Bulletin* 43.4, pp. 887–908.

Moore, B. C. (1996). “Perceptual Consequences of Cochlear Hearing Loss and their Implications for the Design of Hearing Aids”. In: *Ear and Hearing* 17.2, pp. 133–161.

Mortensen, L., A. S. Meyer, and G. W. Humphreys (2006). “Age-related effects on speech production: A review”. In: *Language and Cognitive Processes* 21.1-3, pp. 238–290.

Nachtegaal, J., D. J. Kuik, J. R. Anema, S. T. Goverts, J. M. Festen, and S. E. Kramer (2009). “Hearing status, need for recovery after work, and psychosocial work characteristics: Results from an internet-based national survey on hearing”. In: *International Journal of Audiology* 48.10, pp. 684–691.

Nielsen, J. B. and T. Dau (2011). “The Danish hearing in noise test”. In: *International Journal of Audiology* 50.3, pp. 202–208.

Norwine, A. C. and O. J. Murphy (1938). “Characteristic Time Intervals in Telephonic Conversation”. In: *Bell System Technical Journal* 17.2, pp. 281–291.

Painton, S. W. (1993). “Objective measure of low-frequency amplification reduction in canal hearing aids with adaptive circuitry.” In: *Journal of the American Academy of Audiology* 4.3, pp. 152–156.

Palmer, A. D., J. T. Newsom, and K. S. Rook (2016). “How does difficulty communicating affect the social relationships of older adults? An exploration using data from a national survey”. In: *Journal of Communication Disorders* 62, pp. 131–146.

Pearsons, K., R. Bennett, and S. Fidell (1977). "Speech levels in various noise environments". In: *Bolt Beranek and Newman* Report No.May, Canoga Park, CA.

Peng, Z. E., L. M. Wangtcb, L. M. Wang, Z. E. Peng, and L. M. Wang (2016). "Effects of noise , reverberation and foreign accent on native and non-native listeners ' performance of English speech comprehension". In: *The Journal of the Acoustical Society of America* 139, pp. 2772–2339.

Pichora-Fuller, M. K. et al. (2016). "Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL)". In: *Ear and Hearing*. Vol. 37, 5S–27S.

Picou, E. M. and T. A. Ricketts (2018). "The relationship between speech recognition, behavioural listening effort, and subjective ratings". In: *International Journal of Audiology* 57.6, pp. 457–467.

Piquado, T., J. I. Benichov, H. Brownell, and A. Wingfield (2012). "The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening". In: *International Journal of Audiology* 51.8, pp. 576–583.

Pouw, W. and J. Holler (2020). "Timing in conversation is dynamically adjusted turn by turn: Evidence for lag-1 negatively autocorrelated turn taking times in telephone conversation". In: *Preprint*.

Powers, T., M. Froehlich, E. Branda, and J. Weber (2018). "Clinical Study Shows Significant Benefit of Own Voice Processing". In: *Hearing Review* 25.2, pp. 30–34.

Riest, C., A. B. Jorschick, and J. P. de Ruiter (2015). "Anticipation in turn-taking: Mechanisms and information sources". In: *Frontiers in Psychology* 6.89.

Roberts, F., P. Margutti, and S. Takano (2011). "Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese". In: *Discourse Processes* 48.5, pp. 331–354.

Roberts, S. G. et al. (2015). "The effects of processing and sequence organization on the timing of turn taking: A corpus study". In: *Frontiers in Psychology* 6, p. 509.

Rönnberg, J., M. Rudner, C. Foo, and T. Lunner (2008). "Cognition counts: A working memory system for ease of language understanding (ELU)". In: *International Journal of Audiology*. Vol. 47. SUPPL. 2.

Rönnberg, J. et al. (2013). "The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances". In: *Frontiers in Systems Neuroscience* 7.JUNE, pp. 1–17.

Sacks, H., E. a. Schegloff, and G. Jefferson (1974). "A simplest systematics for the organization of turn taking for conversation". In: *Language* 50.4, pp. 696–735.

Sarampalis, A., S. Kalluri, B. Edwards, and E. Hafter (2009). "Objective measures of listening effort: Effects of background noise and noise reduction". In: *Journal of Speech, Language, and Hearing Research* 52.5, pp. 1230–1240.

Schegloff, E. A. (2000). "Overlapping talk and the organization of turn-taking for conversation". In: *Language in Society* 29.01, pp. 1–63.

Schneider, B. A. and S. Parker (1990). "Intensity discrimination and loudness for tones in broadband noise". In: *Perception and Psychophysics* 47.1, pp. 92–94.

Schober, M. F. and H. H. Clark (1989). "Understanding by addressees and overhearers". In: *Cognitive Psychology* 21.2, pp. 211–232.

Segaert, K., L. Menenti, K. Weber, K. M. Petersson, and P. Hagoort (2012). "Shared syntax in language production and language comprehension-An fMRI study". In: *Cerebral Cortex* 22.7, pp. 1662–1670.

Sjerps, M. J. and A. S. Meyer (2015). "Variation in dual-task performance reveals late initiation of speech planning in turn-taking". In: *Cognition* 136, pp. 304–324.

Sørensen, A. J. M., M. Fereczkowski, and E. N. MacDonald (in press). "The effects of noise and second language on conversational dynamics in task dialogue". In: *Accepted for publication in Trends in Hearing.*

Sørensen, A. J. M. (2021). *PDF and PSD files of DiapixDK picture materials - adapted Danish version.*

Sørensen, A. J. M. and W. O. Brimijoin (2020). "The development of a multimodal communication behaviour capture system". In: *12th Speech in Noise Workshop, Toulouse, France, January 2020.*

Sørensen, A. J. M., M. Fereczkowski, and E. N MacDonald (2020a). "Effects of noise and L2 on the timing of turn taking in conversation". In: *Proceedings of the International Symposium on Auditory and Audiological Research* 7.August, pp. 85–92.

Sørensen, A. J. M., M. Fereczkowski, and E. N. MacDonald (2018). *Task dialog by native-Danish talkers in Danish and English in both quiet and noise.*

Sørensen, A. J. M., E. N. MacDonald, and T. Lunner (2021, submitted). "Conversational dynamics in task dialogue between normal-hearing and hearing-impaired interlocutors". In: *Manuscript submitted for publication.*

Sørensen, A. J. M., E. N. Macdonald, and T. Lunner (2020b). "Timing of turn taking between normal-hearing and hearing-impaired interlocutors". In: *Proceedings of the International Symposium on Auditory and Audiological Research* 7.7, pp. 37–44.

Stephens, D. and F. Zhao (1996). "Hearing impairment: Special needs of the elderly". In: *Folia Phoniatrica et Logopaedica* 48.3, pp. 137–142.

Stivers, T. et al. (2009). "Universals and cultural variation in turn-taking in conversation". In: *Proceedings of the National Academy of Sciences of the United States of America* 106.26, pp. 10587–10592.

Strauss, D. J. et al. (2010). "Electrophysiological correlates of listening effort: Neurodynamical modeling and measurement". In: *Cognitive Neurodynamics* 4.2, pp. 119–131.

Ten Bosch, L., N. Oostdijk, and L. Boves (2005). "On temporal aspects of turn taking in conversational dialogues". In: *Speech Communication* 47.1-2, pp. 80–86.

Torreira, F., S. Bögels, and S. C. Levinson (2015). "Breathing for answering: The time course of response planning in conversation". In: *Frontiers in Psychology* 6.MAR, pp. 1–11.

Tuomainen, O., V. Hazan, and L. Taschenberger (2019). "Speech communication in background noise: Effects of aging". In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Ed. by M. T. Sasha Calhoun, Paola Escudero and P. Warren. Canberra, Australia: Australasian Speech Science and Technology Association Inc., pp. 805–809.

Van Engen, K. J., M. Baese-Berk, R. E. Baker, A. Choi, M. Kim, and A. R. Bradlow (2010). "The Wildcat Corpus of native- and foreign-accented English: communicative efficiency across conversational dyads with varying language alignment profiles." In: *Language and Speech* 53.Pt 4, pp. 510–40.

Watson, S., A. J. M. Sørensen, and E. N. MacDonald (2020). "The effect of conversational task on turn taking in dialogue". In: *Proceedings of ISAAR 2019: Auditory Learning in Biological and Artificial Systems. 7th International Symposium on Auditory and Audiological Research. Nyborg, Denmark*. Vol. 43. August, pp. 61–68.

Weisser, A. and J. M. Buchholz (2019). "Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions". In: *The Journal of the Acoustical Society of America* 145.1, pp. 349–360.

Wendt, D., R. K. Hietkamp, and T. Lunner (2017). “Impact of noise and noise reduction on processing effort: A pupillometry study”. In: *Ear and Hearing* 38.6, pp. 690–700.

Wester, M., M. L. García Lecumberri, and M. Cooke (2014). “DIAPIX-FL: A symmetric corpus of problem-solving dialogues in first and second languages”. In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* 2.September, pp. 509–513.

Wheeldon, L. R. and W. J. Levelt (1995). “Monitoring the Time Course of Phonological Encoding”. In: *Journal of Memory and Language* 34.3, pp. 311–334.

Wijngaarden, S. J. V., H. J. M. Steeneken, T. Houtgast, S. J. V. Wijngaarden, H. J. M. Steeneken, and T. Houtgast (2002). “Quantifying the intelligibility of speech in noise for non-native listeners”. In: *The Journal of the Acoustical Society of America* 111.4.

Wilson, M. and T. P. Wilson (2005). “An oscillator model of the timing of turn-taking”. In: *Psychonomic Bulletin and Review* 12.6, pp. 957–968.

Wilson, R. H., R. A. McArdle, and S. L. Smith (2007). “An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss”. In: *Journal of Speech, Language, and Hearing Research* 50.4, pp. 844–856.

Zyl, M. van and J. J. Hanekom (2011). “Speech perception in noise: a comparison between sentence and prosody recognition”. In: *Journal of Hearing Science* 1.2, pp. 54–56.

Contributions to Hearing Research

Vol. 1: *Gilles Pigasse*, Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, 2008.
External examiners: Mark Lutman, Stefan Stenfeld

Vol. 2: *Olaf Strelcyk*, Peripheral auditory processing and speech reception in impaired hearing, 2009.
External examiners: Brian Moore, Kathrin Krumbholz

Vol. 3: *Eric R. Thompson*, Characterizing binaural processing of amplitude-modulated sounds, 2009.
External examiners: Michael Akeroyd, Armin Kohlrausch

Vol. 4: *Tobias Piechowiak*, Spectro-temporal analysis of complex sounds in the human auditory system, 2009.
External examiners: Jesko Verhey, Steven van de Par

Vol. 5: *Jens Bo Nielsen*, Assessment of speech intelligibility in background noise and reverberation, 2009.
External examiners: Björn Hagerman, Ejnar Laukli

Vol. 6: *Helen Connor*, Hearing aid amplification at soft input levels, 2010.
External examiners: Inga Holube, Birgitta Larsby

Vol. 7: *Morten Løve Jepsen*, Modeling auditory processing and speech perception in hearing-impaired listeners, 2010.
External examiners: Birger Kollmeier, Ray Meddis

Vol. 8: *Sarah Verhulst*, Characterizing and modeling dynamic processes in the cochlea using otoacoustic emissions, 2010.
External examiners: David Kemp, Stephen Neely

Vol. 9: *Sylvain Favrot*, A loudspeaker-based room auralization system for auditory research, 2010.
External examiners: Bernhard Seeber, Michael Vorländer

Vol. 10: *Sébastien Santurette*, Neural coding and perception of pitch in the normal and impaired human auditory system, 2011.
External examiners: Christopher Plack, Christian Lorenzi

Vol. 11: *Iris Arweiler*, Processing of spatial sounds in the impaired auditory system, 2011.
External examiners: Joost Festen, Jürgen Tchorz

Vol. 12: *Filip Munch Rønne*, Modeling auditory evoked potentials to complex stimuli, 2012.
External examiners: Bob Burkard, Stephen Neely

Vol. 13: *Claus Forup Corlin Jespersgaard*, Listening in adverse conditions: Masking release and effects of hearing loss, 2012.
External examiners: Stuart Rosen, Christian Lorenzi

Vol. 14: *Rémi Decorsière*, Spectrogram inversion and potential applications for hearing research, 2013.
External examiners: Michael Stone, Oded Ghitza

Vol. 15: *Søren Jørgensen*, Modeling speech intelligibility based on the signal-to-noise envelope power ration, 2014.
External examiners: John Culling, Martin Cooke

Vol. 16: *Kasper Eskelund*, Electrophysiological assessment of audiovisual integration in speech perception, 2014.
External examiners: Lawrence Rosenblum, Matthias Gondan

Vol. 17: *Simon Krogholt Christiansen*, The role of temporal coherence in auditory stream segregation, 2014.
External examiners: Shihab Shamma, Guy Brown

Vol. 18: *Márton Marschall*, Capturing and reproducing realistic acoustic scenes for hearing research, 2014.
External examiners: Sascha Spors, Ville Pulkki

Vol. 19: *Jasmina Catic*, Human sound externalization in reverberant environments, 2014.
External examiners: Bernhard Seeber, Steven van de Par

Vol. 20: *Michał Feręczkowski*, Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.
External examiners: Christopher Plack, Enrique Lopez-Poveda

Vol. 21: *Alexandre Chabot-Leclerc*, Computational modeling of speech intelligibility in adverse conditions, 2015.
External examiners: Steven van de Par, John Culling

Vol. 22: *Federica Bianchi*, Pitch representations in the impaired auditory system and implications for music perception, 2016.
External examiners: Ingrid Johnsrude, Christian Lorenzi

Vol. 23: *Johannes Zaar*, Measures and computational models of microscopic speech perception, 2016.
External examiners: Judy Dubno, Martin Cooke

Vol. 24: *Johannes Käsbach*, Characterizing apparent source width perception, 2016.
External examiners: William Whitmer, Jürgen Tchorz

Vol. 25: *Gusztáv Löcsei*, Lateralized speech perception with normal and impaired hearing, 2016.
External examiners: Thomas Brand, Armin Kohlrausch

Vol. 26: *Suyash Narendra Joshi*, Modelling auditory nerve responses to electrical stimulation, 2017.
External examiners: Laurel Carney, Bob Carlyon

Vol. 27: *Henrik Gerd Hassager*, Characterizing perceptual externalization in listeners with normal, impaired and aided-impaired hearing, 2017.
External examiners: Volker Hohmann, Piotr Majdak

Vol. 28: *Richard Ian McWalter*, Analysis of the auditory system via synthesis of natural sounds, speech and music, 2017.
External examiners: Maria Chait, Christian Lorenzi

Vol. 29: *Jens Cubick*, Characterizing the auditory cues for the processing and perception of spatial sounds, 2017.
External examiners: Ville Pulkki, Pavel Zahorik

Vol. 30: *Gerard Encina-Llamas*, Characterizing cochlear hearing impairment using advanced electrophysiological methods, 2017.
External examiners: Roland Schaette, Ian Bruce

Vol. 31: *Christoph Scheidiger*, Assessing speech intelligibility in hearing-impaired listeners, 2018.
External examiners: Enrique Lopez-Poveda, Tim Jürgens

Vol. 32: *Alan Wiinberg*, Perceptual effects of non-linear hearing aid amplification strategies, 2018.
External examiners: Armin Kohlrausch, James Kates

Vol. 33: *Thomas Bentsen*, Computational speech segregation inspired by principles of auditory processing, 2018.
External examiners: Stefan Bleeck, Jürgen Tchorz

Vol. 34: *François Guérit*, Temporal charge interactions in cochlear implant listeners, 2018.
External examiners: Julie Arenberg, Olivier Macherey

Vol. 35: *Andreu Paredes Gallardo*, Behavioral and objective measures of stream segregation in cochlear implant users, 2018.
External examiners: Christophe Micheyl, Monita Chatterjee

Vol. 36: *Søren Fuglsang*, Characterizing neural mechanisms of attention-driven speech processing, 2019.
External examiners: Shihab Shamma, Maarten de Vos

Vol. 37: *Borys Kowalewski*, Assessing the effects of hearing-aid dynamic-range compression on auditory signal processing and perception, 2019.
External examiners: Brian Moore, Graham Naylor

Vol. 38: *Helia Relaño Iborra*, Predicting speech perception of normal-hearing and hearing-impaired listeners, 2019.
External examiners: Ian Bruce, Armin Kohlrausch

Vol. 39: *Axel Ahrens*, Characterizing auditory and audio-visual perception in virtual environments, 2019.
External examiners: Pavel Zahorik, Piotr Majdak

Vol. 40: *Niclas A. Janssen*, Binaural streaming in cochlear implant patients, 2019.
External examiners: Tim Jürgens, Hamish Innes-Brown

Vol. 41: *Wiebke Lamping*, Improving cochlear implant performance through psychophysical measures, 2019.
External examiners: Dan Gnasia, David Landsberger

Vol. 42: *Antoine Favre-Félix*, Controlling a hearing aid with electrically assessed eye gaze, 2020.
External examiners: Jürgen Tchorz, Graham Naylor

Vol. 43: *Raul Sanchez-Lopez*, Clinical auditory profiling and profile-based hearing-aid fitting, 2020.
External examiners: Judy R. Dubno, Pamela E. Souza

Vol. 44: *Juan Camilo Gil Carvajal*, Modeling audiovisual speech perception , 2020.
External examiners: Salvador Soto-Faraco, Kaisa Maria Tiippana

Vol. 45 *Charlotte Amalie Emdal Navntoft*, Improving cochlear implant performance with new pulse shapes: a multidisciplinary approach, 2020.
External examiners: Andrej Kral, Johannes Frijns

Vol. 46 *Naim Mansour*, Assessing hearing device benefit using virtual sound environments, 2021.
External examiners: Virginia Best, Pavel Zahorik

Vol. 47 *Anna Josefine Munch Sørensen*, The effects of noise and hearing loss on conversational dynamics, 2021.
External examiners: Martin Cooke, William Whitmer

The end.

To be continued…

Participating in conversation is an integral part of human social interaction, and having a hearing impairment makes it difficult to communicate, especially in noisy environments. This can lead to hearing-impaired individuals withdrawing from social interactions, eventually leading to social isolation. One of the most important outcomes of hearing rehabilitation, therefore, is to regain people's ability to partake in social interactions, but we often have to generalize results from tests that only involve one part of communication, either comprehension or production. Conversation, however, involves an overlap between the two, and it is a dynamic feedback process between interlocutors who may adapt to each other's difficulties to alleviate communication barriers. This thesis aimed to find objective measures of conversational dynamics that varied in a systematic way when people experienced increased communication difficulty. In four experiments, we investigated the effects of noise on conversational dynamics between normal-hearing pairs of interlocutors, and between normal-hearing/hearing-impaired pairs either seated separately or face-to-face. When communicating in noise, hearing-impaired individuals were less precise in their timing of turn-taking, they produced longer utterances, and they spoke slower and louder. When receiving compensation for their hearing loss, hearing-impaired individuals were able to time their responses with higher precision, and they spoke faster and produced shorter utterances. The results are promising for the prospect of using conversational dynamics as objective measures for evaluating the performance of different hearing assistive devices' processing strategies and features.

DTU Health Tech
Department of Health Technology

Ørsteds Plads
Building 352
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 39 50
www.dtu.dk