UNIVERSITY OF HELSINKI

Asymmetrical Lombard Effect

Conversating in Loud and Quiet Environments Simultaneously

Alexandra Wikström
Master's thesis
Linguistic Diversity and Digital Humanities
Phonetics
Faculty of Arts
University of Helsinki
May 2022

Tiivistelmä

Tiedekunta: Humanistinen tiedekunta
Koulutusohjelma: Kielellisen diversiteetin ja digitaalisten ihmistieteiden maisteriohjelma
Opintosuunta: Fonetiikka
Tekijä: Alexandra Wikström
Työn nimi: Asymmetrinen Lombard-efekti – Yhtäaikainen keskustelu meluisassa ja hiljaisessa ympäristössä
Työn laji: Maisterintutkielma
Kuukausi ja vuosi: Toukokuu 2022
Sivumäärä: 36
Avainsanat: Fonetiikka, Lombard-efekti, Lombard-puhe, keskustelupuhe, spontaani puhe
Ohjaaja tai ohjaajat: Juraj Šimko
Säilytyspaikka: Helsingin yliopiston kirjasto / E-thesis
Tiivistelmä:

Ihmiset muuttavat äänentuotantoaan kuuluvammaksi meluisassa ympäristössä refleksinomaisesti. Tätä ilmiötä kutsutaan Lombard-efektiksi. Efekti saa puhujan tuottamaan Lombard-puhetta, jota on tutkittu jo yli vuosisadan ajan eri näkökulmista. Lombard-puheen akustiikalle ominaista ovat korotettu äänenpainetaso, korotettu puheäänen perustaajuus, muutokset äänen osataajuuksissa sekä muissa äänen spektrin rakenteissa. Lisäksi Lombard-puheessa vokaalien pituuksilla on tapana kasvaa, ja äärimmäisissä meluolosuhteissa hyperartikulaatiota voi esiintyä. Puhetilanteeseen sisältyvä kommunikatiivinen aspekti on keskeistä ilmiön synnylle.

Tämän tutkielman tavoitteena oli tutkia puheentuottoa keskustelutilanteessa, jossa samanaikaisesti toinen keskustelijoista on altistettuna melulle ja tuottaa täten Lombard-puhetta, ja toinen keskustelija kommunikoi hiljaisuudessa ilman taustamelun suoria vaikutuksia, ja selvittää, onko puheen akustiikassa tai ymmärrettävyydessä eroavaisuuksia tällaisessa epäsymmetrisessä tilanteessa verrattuna symmetriseen puhetilanteeseen, jossa molempien puhujien ääniympäristö on sama. Tutkimusta varten kaksi paria suomenkielisiä keskustelijoita (yhteensä neljä osallistujaa, kaikki naisia) ratkoivat pareittain sudokupohjaisia tehtäviä kolmessa eri taustamelutilanteessa: (1) hiljaisuudessa, (2) molempien ollessa taustamelussa (symmetrinen), ja (3) vain toisen keskustelijan ollessa taustamelussa (asymmetrinen). Taustamelu, jota soitettiin koehenkilöille 75 dB äänenpainetasolla, oli laadultaan cocktail-melua, joka sisältää niin kutsuttua puheensorinaa jossa useampi puhuja puhuu päällekkäin. Keskustelut äänitettiin ja niistä kerättiin yhteensä 453 maalitavua, joista kaikista analysoitiin keskimääräinen äänenpainetaso, ja 417 maalitavusta analysoitiin keskimääräinen perustaajuus. Äänenpainetason ja perustaajuuden arvot normalisoitiin ja arvoille suoritettiin keskiarvoja ja variansseja vertailevat tilastolliset testit.

Odotetusti kaikki puhujat korottivat äänenpainetasoaan ja perustaajuuttaan siirryttäessä hiljaisesta keskustelutilanteesta symmetriseen taustamelutilanteeseen, jossa molemmat keskustelukumppanit tuottivat Lombard-puhetta. Henkilöt, jotka asymmetrisessä keskustelutilanteessa olivat itse hiljaisuudessa ja kommunikoivat keskustelukumppanille, joka oli melussa, korottivat sekä äänenpainetasoaan että perustaajuuttaan asymmetrisessä keskustelutilanteessa verrattuna hiljaiseen keskustelutilanteeseen. Lisäksi toinen näistä puhujista korotti sekä äänenpainetasoaan että perustaajuuttaan lähes oman Lombard-puheensa tasolle, jota mitattiin symmetrisessä tilanteessa. Puhujat, jotka olivat altistettuna melulle asymmetrisessä tilanteessa, käyttivät keskimäärin matalampaa äänenpainetasoa asymmetrisessä kuin symmetrisessä tilanteessa, vaikka tuottivatkin Lombard-puhetta molemmissa tilanteissa. Väärin kuultuja maalitavuja ei havaittu asymmetrisessä tilanteessa, vaan henkilöt, jotka olivat kyseisessä tilanteessa hiljaisuudessa, onnistuivat korottamaan ääntään tarvittavalle tasolle, jotta ratkaiseva tieto saatiin kommunikoitua melussa olevalle henkilölle.

Tämä tutkimus osoitti, että kahden keskustelukumppanin ääniympäristöjen ollessa eriävät, kumpikaan keskustelijoista ei tuota täysin sentyyppistä puhetta, joka olisi sopivaa heidän senhetkiseen ääniympäristöönsä, vaan puheentuottoon vaikuttaa myös välillisesti keskustelukumppanin ääniympäristö. Lisäksi tutkimus osoitti, että siinä missä puhetilanteen kommunikatiivisuus voi lisätä Lombard-efektin vaikutuksia, se voi myös häivyttää niitä. Jatkotutkimuksissa tulisi kerätä enemmän dataa ja suorittaa datalle laajempaa analyysiä.

Abstract

Faculty: Faculty of Arts
Degree programme: Master's programme in Linguistic Diversity and Digital Humanities
Study track: Phonetics
Author: Alexandra Wikström
Thesis title: Asymmetrical Lombard Effect – Conversating in Loud and Quiet Environments Simultaneously
Level: Master's thesis
Month and year: May 2022
Number of pages: 36
Keywords: Phonetics, Lombard effect, Lombard speech, conversational speech, spontaneous speech
Supervisors: Juraj Šimko
Where deposited: Helsinki University Library – Helda / E-thesis
Abstract:
Humans increase their vocal efforts in a noisy environment in a reflex-like manner. This phenomenon is called the Lombard effect. The effect causes the speaker to produce Lombard speech, which has been researched for over a century from different standpoints. Lombard speech is characterized by increased mean energy intensity level, increased fundamental frequency, changes in the formant frequencies, and in other spectral qualities of the voice. In addition, vowel durations tend to increase and in extreme noise conditions, a speaker might hyperarticulate. The communicative aspect of a speech situation is essential to the emergence of the phenomenon.

The goal of this thesis was to examine speech production in a conversational situation where simultaneously one of the interlocutors engaged in a conversation is subjected to noise and is thus producing Lombard speech, while the other interlocutor is communicating in silence without the direct effects of background noise, and to determine, whether there are differences in the acoustics or the intelligibility of speech in such an asymmetrical speech situation compared to a symmetrical situation where the noise environment of the interlocutors is the same. Two pairs of Finnish speakers (4 participants altogether, all female) were recorded doing sudoku-based tasks in three different background noise conditions: (1) in quiet, (2) with both interlocutors in noise (symmetrical), and (3) with only one of the interlocutors subjected to noise (asymmetrical). The background noise, played at 75 dB, was cocktail noise, which includes unintelligible speech from simultaneous speakers. Altogether 453 target syllables were collected, and the mean energy intensity level was extracted from each syllable. Mean fundamental frequency (f_0) data was extracted from 417 target syllables. The values of f_0 and intensity were normalized and statistical tests comparing means and variances were carried out on the data.

Expectedly all participants increased their intensity level and f_0 from the quiet to the symmetrical condition, where both interlocutors produced Lombard speech. The participants who during the asymmetrical condition were in silence and communicated to the interlocutor who was in noise increased both their intensity and f_0 in the asymmetrical condition compared to the quiet condition. In addition, one of these participants increased both measures to nearly the levels that were measured from her Lombard speech in the symmetrical condition. The participants who were subjected to noise during the asymmetrical condition on average used lower intensity levels in the asymmetrical condition than in the symmetrical condition, even though they produced Lombard speech during both. No target syllables were misheard during the asymmetrical condition, rather, the participants who were in silence during said condition managed to increase their vocal efforts to a level that ensured the communication of crucial information to the person in noise.

This experiment demonstrated that when the sound environments of two interlocutors are different, neither of the interlocutors produces speech that would be completely suitable for their respective environments but are indirectly affected by the sound environments of their conversational partners. In addition, it was shown that while communicativeness can increase the effects of the Lombard effect, it can also decrease them. For further research into the topic more data should be gathered, and wider analyses should be carried out.

TABLE OF CONTENTS

1 Introduction 1

2 Lombard speech 2

2.1 Lombard speech qualities 2

2.1.1 Acoustics 2

2.1.2 Articulation and phonation 3

2.2 Comparisons with shouted speech 4

2.3 Communicative aspect of Lombard speech 5

3 Research question and hypothesis 6

4 Experiment 7

4.1 Methods 9

4.2 Analysis 10

4.3 Results 11

4.3.1 Temporal variations of intensity 11

4.3.2 Intensity variations of groups 15

4.3.3 Individual variations of intensity 17

4.3.4 Temporal variations of fundamental frequency 19

4.3.5 Fundamental frequency variations of groups 22

4.3.6 Individual variations of fundamental frequency 24

4.4 Discussion 27

5 Conclusions 30

References 32

1 Introduction

The phenomenon of increasing one's vocal efforts in the presence of background noise, i.e., the Lombard effect, has been investigated for over a century ever since it was first discovered by Étienne Lombard in 1911 (Lombard, 1911). The research has since then spread from acoustic and articulatory studies of humans (e.g., Garnier & Heinrich, 2013; Šimko et al., 2014; Van Heusden et al., 1979) to studies of the Lombard effect in animals such as domestic cats (Nonaka et al., 1997) and tree swallows (Leonard & Horn, 2005), and more recently into the field of speech synthesis (e.g., Marxer et al., 2018; Šimko et al., 2020) and many others.

Though the importance of the communicative aspect in the emergence of the effect has been acknowledged before (Lu & Cooke, 2009a), it is only in the last decade or so that it has been taken to be a serious aspect of data collection in Lombard speech research (Garnier et al. 2010; Garnier & Heinrich, 2013). More specifically, the research into one-sided, asymmetrical Lombard speech situations in a conversational setting seems to be lacking entirely. This refers to situations where one of two interlocutors in a conversation is subjected to noise, while the other one is not. This might be a familiar phenomenon from everyday life, for example in a situation where two people are talking on the phone and one of the interlocutors is in a café, a bar, or another noisy environment. This can also happen in face-to-face situations where for example the source of the noise (e.g., a machine) is close to one of the interlocutors but not to the other, resulting in the interlocutor next to the noise source having more issues in comprehending the other interlocutor further from the noise.

Compared to a more normal conversational situation in noise that induces the Lombard effect in both interlocutors, the described asymmetrical situation only evokes Lombard speech in the interlocutor who is subject to the noise. This leaves the interlocutor not affected by noise with the task of increasing vocal efforts in a way that makes them intelligible without having the innate sense of what the appropriate level of vocal effort might be compared to the noise level. Thus, this interlocutor must rely on the feedback of their conversational partner to know if the speech is intelligible enough compared to the background noise and might resort to a form of shouting to ensure it. As the research into speech production during such an asymmetrical setting appears not to have been done previously, it will be the goal of this paper.

In this thesis, I will go through the various qualities of Lombard speech, a brief comparison of how Lombard speech differs from otherwise loud speech, and what the importance of a

communicative situation is to the phenomenon and its research. After that, I will present the research question and hypothesis on which the experiment carried out for this paper was based, and lastly, lay out the proceedings of the study and discuss its results and implications for the future.

2 Lombard speech

2.1 Lombard speech qualities

2.1.1 Acoustics

Lombard speech, that is, speech produced in the presence of background noise, has qualities that make it clearly stand out from speech produced without the effects of background noise. The most prominent feature of Lombard speech is an approximated increase of 3 dB in speech intensity level when the background noise intensity level is increased by 10 dB (conversational speech, ambient noise background; Van Heusden et al., 1979). In Van Heusden's study it was determined that the sound intensity level required for the background noise to induce the effect, is around 40 dB. This is in line with previous and later research, with reported intensity levels of 35–45 dB of background noise needed (Bottalico et al., 2017).

In addition to the increases in the intensity levels of speech, the Lombard effect is linked to increases in the fundamental frequency (f_0), increases in the first formant F1, shifts in the second formant F2, shifting of the spectral center of gravity (CoG) and flattening of the spectral tilt (Junqua, 1993; Lu & Cooke, 2008). The CoG is known to rise more for females than males, which is one of the few differences in Lombard speech between genders. It refers to the frequency area that has the most energy concentrated on it. Females have increased energy between 4 and 5 kHz and males between 2 and 4 kHz (Garnier & Henrich, 2013; Junqua, 1993). Females also tend to increase F1 more than males (Garnier & Henrich 2013). F2 is known to shift either upwards or downwards (Lu & Cooke, 2008; Uemura et al., 2010), which is understandable as F2 varies together with the forward and backward motion of the tongue, and the F2 shifts in Lombard speech might thus represent the exaggerated articulation of certain speech sounds. Some acoustic changes of the Lombard effect, specifically ones in f_0, F1, and CoG, are regarded to be the direct result of increased sound pressure levels of louder speech (Garnier & Henrich, 2013) and can be seen as not necessarily being direct contributors to

heightened intelligibility of Lombard speech but rather a side effect of the strategies that are utilized to maintain intelligibility (Lu & Cooke, 2009b). As a counterpoint to this argument though, it needs to be pointed out that increases in f_0 also happen when speech is more enunciated, and such changes in prosody are connected to the nature of increased intelligibility in Lombard speech.

2.1.2 Articulation and phonation

As mentioned earlier, some acoustic changes, like the ones in f_0 and F1, are in part the result of articulatory and phonatory changes caused by the Lombard effect. f_0 is affected by a rise in subglottal pressure and F1 rises with the widening of the jaw, both due to increased vocal efforts. At a background noise level of 80 dB, the movement of articulators like the lower jaw is known to increase and leads to hyperarticulation, i.e., exaggerated articulation (Šimko et al., 2014). While the trajectories of articulators get longer, more time is needed to complete an articulatory movement if a similar speed is maintained, or the speed of the articulators is increased to maintain a certain speaking rate. As Šimko et al. (2016) demonstrated, the hyperarticulatory movements of the jaw are also strongly linked to hyperarticulatory movements of the tongue. The hyperarticulatory movements of the tongue, however, are not as drastic as the hyperarticulatory movements of the jaw. Hyperarticulation brought on by the Lombard effect can also appear in lip movements. Depending on the speech sound, the lips might become more protruded, wider, or have an increased aperture compared to non-hyperarticulated speech. All of these features are more strongly prevalent when a visual connection to the other interlocutor exists and can be utilized to increase intelligibility (Fitzpatrick et al., 2015; Garnier et al., 2018).

As an addition to the hyperarticulation affecting durations in speech, speakers also start to emphasize certain parts of speech more in noisy conditions in order to maintain intelligibility. These together lead to higher vowel-to-consonant ratios, i.e., longer durations in vowels and sometimes shortened consonants (Garnier & Henrich 2013; Junqua, 1993). In Lu & Cooke (2009b) it was suggested that the elongated durations of speech sounds would be one of the key qualities of why Lombard speech has such heightened intelligibility compared to speech produced in quiet.

2.2 Comparisons with shouted speech

Much like Lombard speech, shouted speech can clearly be told apart from speech produced in a quiet environment in a "normal" manner. The most obvious quality of shouted speech is the increase in intensity or sound pressure level. The trajectories of the articulators are also increased compared to speech that is not shouted (Xue et al., 2021). With the heightened subglottal pressure which creates louder phonation, the fundamental frequency of shouted speech is increased. To maintain the high subglottal pressure used in shouting, people also tend to utilize higher lung volumes, which in turn increases the air intake into the lungs (Huber et al., 2005).

Although shouted and Lombard speech are both distinguishable from normal speech and share many qualities, some differences set them apart. Whereas Lombard speech is loud speech brought on by the presence of background noise which forces the speaker to increase their vocal efforts in a reflex-like way (though not a reflex as Brumm & Zollinger (2011) point out), shouting uninvoked by the Lombard effect is brought on by higher-level demands, such as distance between the speaker and the listener, or other situations where the speaker feels their vocal efforts need to be increased in order to be understood by others. As demonstrated by Raitio et al. (2013), if asked to shout, people produce different levels of loud speech due to differing concepts of what shouting is. Even though the levels of vocal effort differ between speakers when shouting, the common denominator is the change in the perceived loudness of speech compared to a baseline of "normal" speech fitted for the current environment. Perceived loudness is not only affected by changes in intensity but by changes in parameters like voice quality and f_0 as well (Dawson et al., 2017; Yanushevskaya et al., 2013).

In addition to the difference in what evokes the two speaking styles, there are distinctions in the acoustics of them as well. Whereas Lombard speech is known to have more varied f_0 contours compared to normal speech, in shouted speech the f_0 contours show less variation. Lombard speech has more highs and lows in its f_0 contour whereas shouted speech tends to rise to a high f_0 and stay there continuously (Rostolland, 1982a). The formant qualities of the two speaking styles vary as well. Extreme shouting can lead to speech sounds becoming more similar to one another as the articulatory trajectories are simplified, resulting in the first two formants assimilating with each other (Rostolland, 1982b). In Lombard speech the opposite happens, as vowels usually become more defined and increases in F1 happen (Lu & Cooke, 2008).

Lastly, though both Lombard speech and shouting require higher lung volumes to maintain heightened subglottal pressure, Huber et al. (2005) did discover differences between how the pressure is maintained. While asking people to shout makes people use either increased recoil pressure (increased lung volume) or tenser expiratory muscles to maintain sufficient subglottal pressure depending on the exact instruction, in producing Lombard speech speakers tend to utilize both of these methods together.

2.3 Communicative aspect of Lombard speech

The reason why the effects of background noise on speech production are greater in a communicative situation is simple: Lombard speech (or any other vocalization made under the Lombard effect, at that) is based on the goal of being understood in a noisy environment by another interlocutor. Therefore, making someone read aloud in a laboratory to a nearly non-existent audience without the responsibility of carrying out a conversation successfully does not mimic a normal communicative speaking situation very well. However, there have been differing opinions on how central of a role the communicative aspect plays in the emergence of the effect.

On one hand, a person utilizes a private loop of auditory feedback to control the appropriate level of speech production (Charlip & Burke, 1969). This includes both the feedback of the sound environment the speaker is emerged in, and the side tone, which refers to the speaker hearing their own voice (Natale, 1975). On the other hand, a speaker also adjusts their vocal production in accordance with a public feedback loop (Lane & Tranel, 1971). This refers to the feedback received from the other interlocutor in a conversation, which is where the communicative aspect of the Lombard effect stems from. Lane and Tranel suggested in their 1971 paper that the public loop of auditory feedback is all that a speaker uses and that while not in a communicative situation, a speaker does not care whether they can hear their own speech and would thus not make use of the private auditory feedback loop. Since then, it has been shown that Lombard speech is not only important in a communicative situation, but that people also regulate their speech production via the private loop when not engaged in a conversation (Garnier et al., 2010). In addition, Pick et al. (1989) found that even when consciously trying to suppress the effect while being subjected to noise, speakers are not able to ignore its effects and can only via training start to disregard it.

For decades the speech tasks used in Lombard effect research have been centered around performing monologues and uttering aloud separate prepared sentences or words. There have been clear advantages to using more monologue speech in research: The spontaneous speech used in communication can be problematic because it cannot be controlled as easily as speech in reading tasks to have the same kinds of structures (for example similar syllables) in each noise setting. This leads to difficulties in having uniform data to examine across different conditions, which structured reading tasks do provide.

However, as mentioned before, simple reading tasks largely lack any sort of natural communicative aspects of speech that the Lombard effect is based upon. Studies like Lu & Cooke (2009a) and Garnier et al. (2010) have demonstrated that not only do communicative tasks work better in invoking the Lombard effect rather than situations where the speaker only reads aloud words or sentences, the speech collected from such tasks can indeed be controlled for what kind of phonological structures they contain. This is usually done by incorporating games into the recording sessions which both evoke conversation between the participants and create a controlled vocabulary of words that are used to communicate about the game, which can later be utilized in analyzing the gathered data.

Whether the speech data gathered is rigid read speech or more spontaneously produced, Wagner et al. (2015) bring out an important point: no matter what kind of speech data is gathered, the most important aspect is to acknowledge in what way that speech style might affect the data and investigate it accordingly.

3 Research question and hypothesis

The goal of this research was to investigate conversational speech production whilst both interlocutors are in noise (symmetrical condition), and the situation where only one of the interlocutors is subjected to noise (asymmetrical condition) and also compare these two to conversating in the absence of any background noise. *Are there differences in the intelligibility or the acoustics of speech, when either both interlocutors are in noise, or when just one of the interlocutors is subjected to noise?* This question could help shed light on a few interesting

issues. First of all, the role that the aforementioned public and private loop of auditory feedback play when a person in quiet is trying to communicate to a person in noise. In this situation, the private loop does not offer a reference point for the speaker to adjust their vocal production, but the speaker relies entirely on the public feedback loop of if the interlocutor in noise can understand them. The speech is thus not Lombard speech induced by background noise but needs more conscious efforts to be made intelligible. Secondly, the reflexive nature of the Lombard effect. When a person is subjected to noise but is conversating with a person who is not in noise, the need to actively enhance one's intelligibility is diminished. The person in quiet would hear the other interlocutor even without them producing Lombard speech.

The hypothesis is that during the asymmetrical condition the person in silence will try to increase their vocal efforts even though in a quiet environment, to account for the difficulties in hearing that the interlocutor in noise might have if spoken to with a "normal" voice that suits a quiet environment. Presumably also more misinterpretations and corrections happen in the asymmetrical condition compared to the symmetrical and quiet conditions, due to the interlocutor who is not subjected to noise not having a clear sense as to what the appropriate level of vocal effort would be for the person in noise to hear them. When it comes to the speech production of a person who is subjected to noise in both the symmetrical and the asymmetrical condition, there will likely not be great differences between the two conditions. This is because the person will be producing Lombard speech in both of the conditions, which we know to be a largely involuntary reaction.

4 Experiment

The experimental setting used here was in part modeled after the Fitzpatrick et al. (2015) study on the effect of seeing one's conversational partner during the production of Lombard speech. In that study, a 9x9 sudoku puzzle where the numbers were replaced with syllables was created and used as a task for the participants to ensure the existence of comparable tokens in the speech data while remaining spontaneous. Sudoku puzzles have also been utilized earlier in collecting Lombard speech data, for example in the 2010 study by Cooke and Lu where the effects of informational and energetic masking on Lombard speech production were explored. Cooke and Lu's experimental setting however utilized traditional sudokus which contain only numbers, which limits the initial control that the designer of the task has on the phonological structures that are included in the task and therefore in the subsequent speech data.

For the current study, a similar sudoku task to the one in Fitzpatrick et al. (2015) was created but with syllables that better fit the Finnish setting. The syllables, *kat, kot, kut, kät, köt, tat, tot, tät,* and *töt,* were chosen so that the distinction between speech sounds in their pronunciation might add to the challenge of communicating in noise. All of the syllables start with a plosive sound, either [k] or [t], end in [t], and are followed by the vowels [ɑ], [o], [u], [æ] and [ø]. The included syllables are all combinations that occur in the Finnish language but are not words on their own.

In Fitzpatrick et al. (2015) only pairs who knew each other were recorded. Seen as though studies on interspeaker relations point out that variations in producing conversational speech usually rise from already established social connections (Farley et al., 2013; Pardo et al., 2012), the current experiment utilized only pairs of people who did not previously know each other as to avoid the effects of any possible variation created by previously established social connections.

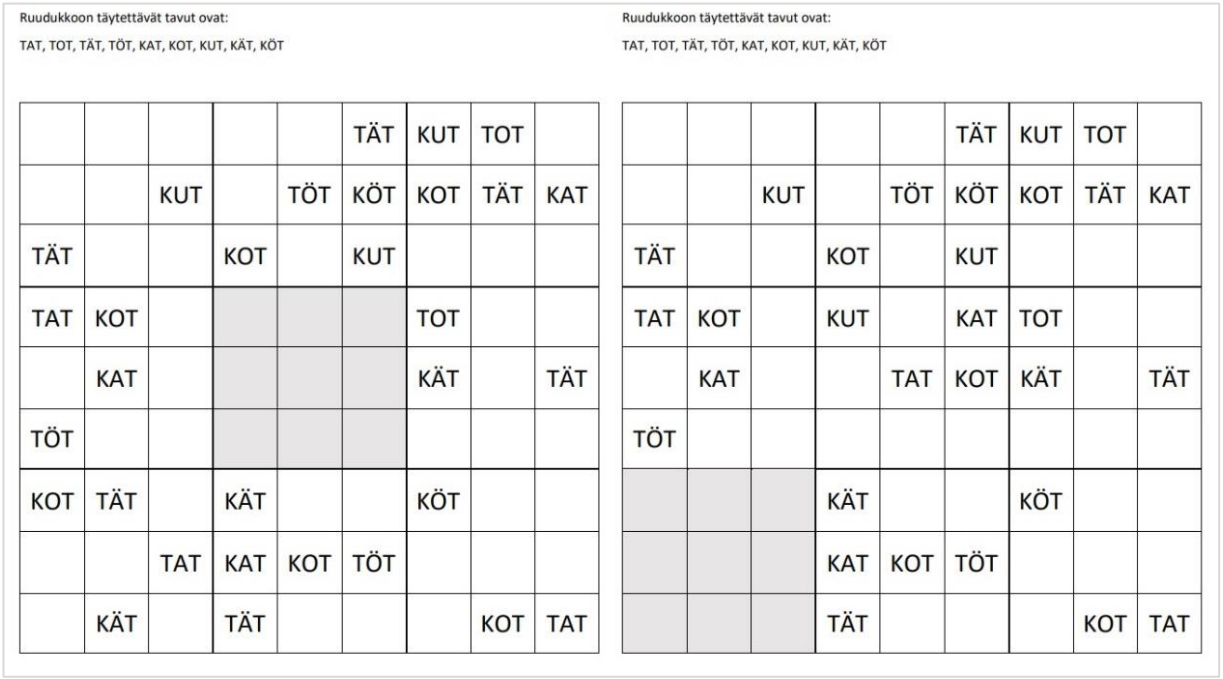


Figure 1. An example of a sudoku utilized in the experiment. The sudoku is the same for both participants but different parts of it are visible to each person, i.e., the grey empty box is in a different part of the sudoku depending on which copy the participant has. The boxes worked as a starting point for the pairs to start communicating on solving the task and the participants were also presented with a list of the possible syllables above the sudokus to help the process of familiarizing themselves with the options that replaced the numbers.

4.1 Methods

Participants were recruited through social media and received a movie ticket as compensation for participation. Four participants, all female and native speakers of Finnish aged between 22 and 27 years, were recorded in pairs while trying to solve the sudoku puzzle as a team. To ensure the need for communication between the participants, a different 3x3 block of the sudoku was hidden from both interlocutors which they needed to fill out with the help of their partner who knew what syllables belonged to that particular part of the sudoku (Figure 1).

The setting was repeated in three different conditions and in the same order for both pairs: (1) in silence, (2) with both interlocutors in noise (symmetrical condition), and (3) with only one of the interlocutors subjected to noise (asymmetrical condition). In the asymmetrical condition the interlocutor who was subjected to noise was always in the right recording booth, and the interlocutor not subjected to noise was always in the left recording booth. Therefore, the data presenting interlocutors who were in noise in both the symmetrical and the asymmetrical condition are referred to as right booth, and the data presenting interlocutors who were in noise during the symmetrical condition but not during the asymmetrical condition are referred to as left booth.

The pairs were instructed to work on the sudoku collaboratively for 15 minutes in the silent condition so as to get familiar with the task and for 10 minutes in the two other conditions. The participants did however quickly grasp how to execute the task during the quiet condition and thus the data from said condition from both sessions was included in the analysis. At the start of each condition a new sudoku was provided for the participants. Neither of the pairs was able to solve a complete sudoku puzzle during a single condition due to the list of syllables not being as intuitive to go through as a list of consecutive numbers.

The background noise used was cocktail noise, meaning a recording of a room with several individuals having simultaneous unintelligible conversations, much like something one would be subjected to in a cocktail party, or another crowded situation. The noise was played at 75 dB to the participants through Sennheiser HD250 Linear II headphones. This level of noise was both loud enough to elicit the Lombard effect but not too loud so as to tire out the participants during the session. The interlocutors were recorded with DPA d:fine 4066 Omnidirectional headset microphones in separate recording booths with visual contact through a window. The

	Quiet		Symmetrical		Asymmetrical	
Syllable	**Left**	**Right**	**Left**	**Right**	**Left**	**Right**
KAT	13	8	9	10	9	9
KOT	9	2	6	5	5	4
KUT	12	13	7	5	6	5
KÄT	10	9	5	4	11	4
KÖT	11	10	5	4	13	11
TAT	17	7	6	8	5	2
TOT	20	20	8	11	8	10
TÄT	14	7	4	6	9	10
TÖT	17	8	4	4	4	10

Table 1. The number of syllables extracted from each of three conditions: quiet, symmetrical, and asymmetrical. "Left" indicates the data from participants who were in silence during the asymmetrical condition and "Right" indicates the data from participants who were subjected to noise during the asymmetrical condition.

participants also received a feed of their own voice and the other interlocutor's voice through the headphones to cancel out the effect of the headphones on the speaker's own feedback loop and the effect of the wall between the participants. A short break was held between the different conditions. The participants were asked for comments about the experiment after the recording session and if they knew what the Lombard effect referred to. None of the participants reported having any prior knowledge of the Lombard effect and gave some comments about the communicational situation they were in that will be discussed at a later point.

4.2 Analysis

Altogether 453 target syllables (the syllables that were a part of the sudoku puzzle) were extracted and analyzed, yielding 3 minutes and 7 seconds of speech. The number of target syllables collected from each condition can be found in Table 1. The distribution of the different syllables is fairly uniform with the majority appearing around 45 to 55 times, with the exception of *tot* appearing 77 times and *kot* only 31 times. Unsurprisingly the quiet condition which went on the longest yielded the most target syllables while the symmetrical condition yielded the least. Annotation of the collected data and the acoustic analyses were carried out with Praat (Boersma & Weenik, 2019). The mean fundamental frequency and the mean-energy intensity level were extracted from each sudoku syllable. For visualization purposes the time point at which each target syllable started was also extracted. 36 data points, mainly from the data

extracted from the quiet condition, had to be excluded from the f_0 analysis because the fundamental frequency could not be calculated, due to lack of clear phonation in creaky or whispering speech.

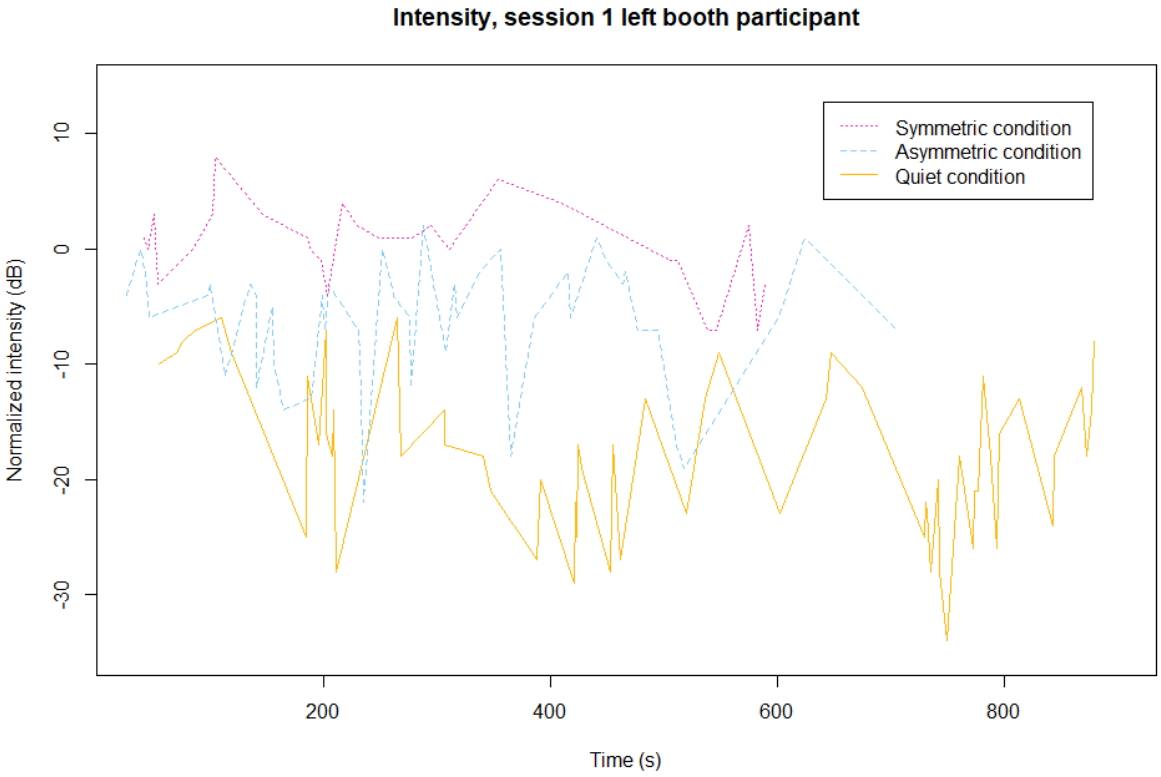
In order to account for differences of range in intensity and frequency between individual speakers, the values of both fundamental frequency and intensity level were normalized. The normalization of values needed to be done by comparing the extracted values to a fitting baseline that would be consistent across speakers. The data from the quiet condition could have been utilized, but due to the Lombard effect creating generally more consistent f_0 and intensity data with less variation than normal speech in silence, the Lombard speech data from the symmetrical condition was chosen as a baseline. The f_0 normalizing was done by converting the difference between the mean f_0 value of an individual speaker in the symmetrical condition and the compared values from Hertz to semitones. Normalization of intensity was carried out correspondingly to display the difference between the mean intensity value of an individual speaker in the symmetrical condition and the measured values. Statistical analysis of the data was carried out with R/RStudio (RStudio Team, 2020) using a confidence interval of 95 percent.

4.3 Results

4.3.1 Temporal variations of intensity

Figures 2 and 3 present the time progression of the intensity data for individual speakers, with the x-axis presenting time in seconds and the y-axis the intensity values. Temporal changes in the data do not show a specific trend of increase or decrease to one certain direction over time, but rather oscillations between lower and higher values throughout the sessions for all participants, presenting the nature of a conversation where the values are the product of different prosodic environments. Some of the larger oscillations are also due to the data points consisting only of the target syllables, which the participants at some points did not produce for longer periods of time.

The quiet condition can be seen yielding the lowest values across speakers, with session one right booth participant (Figure 2) and session two left booth participant (Figure 3) showing



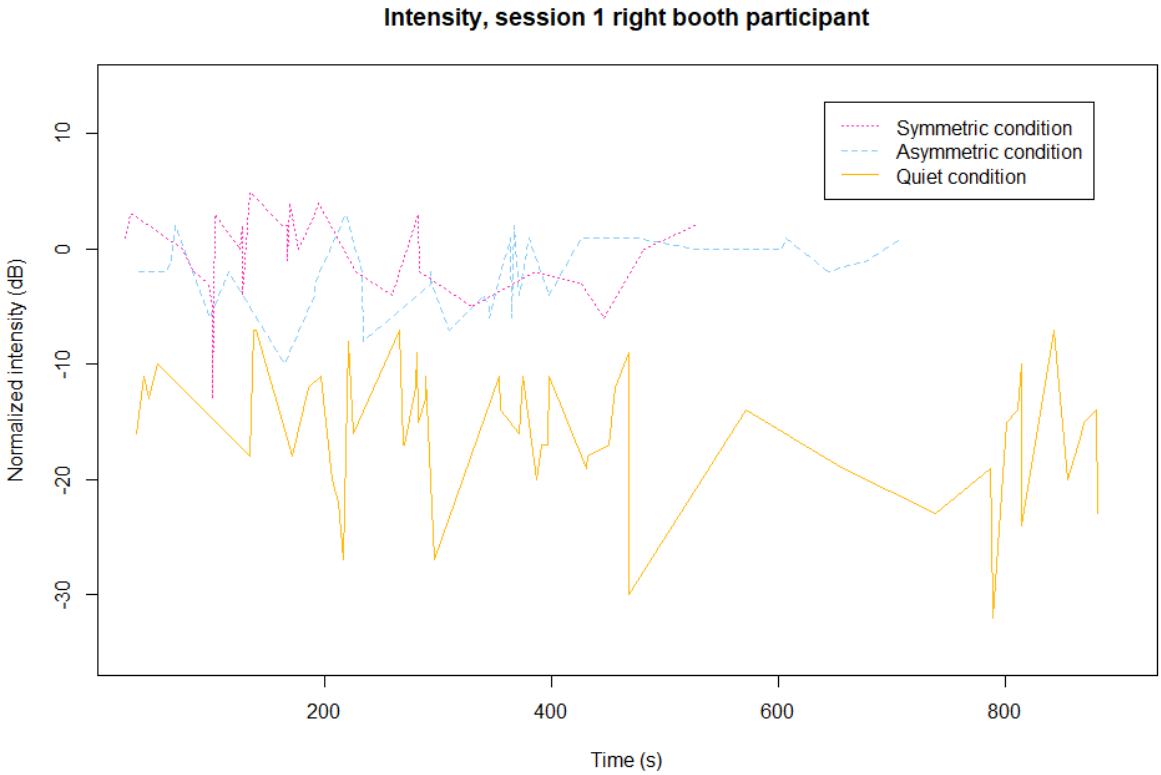
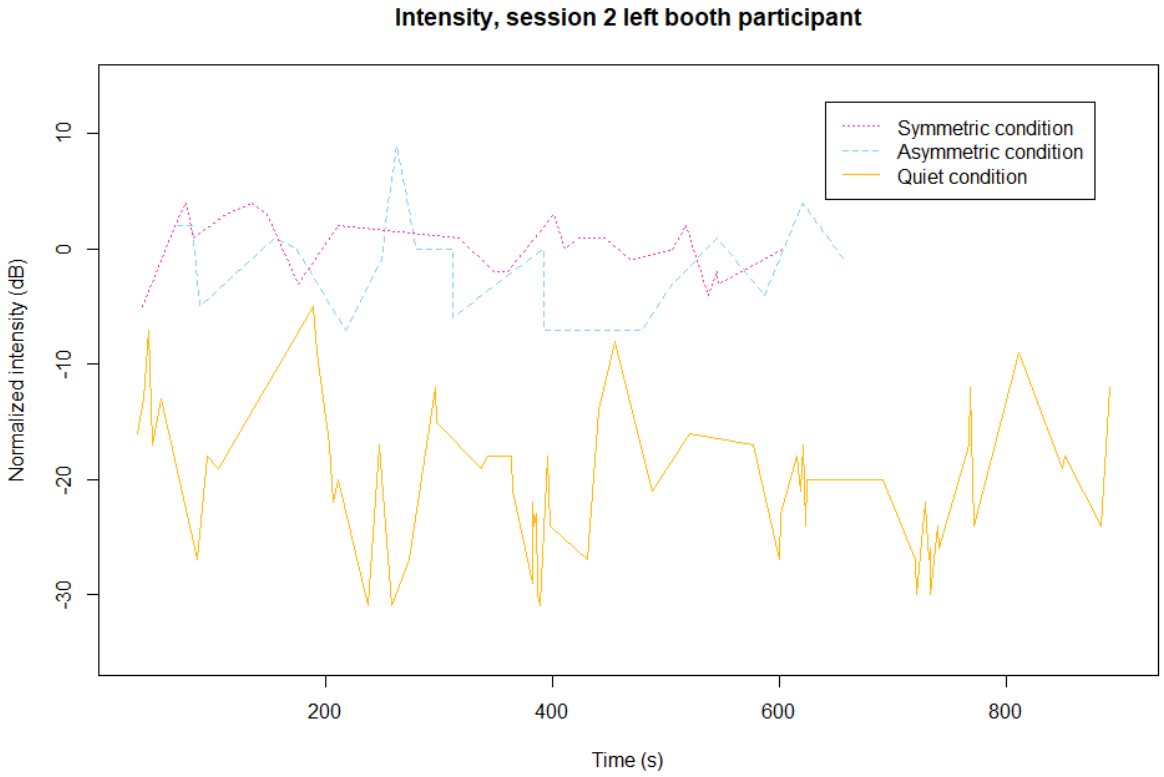


Figure 2. Intensity data time series of target syllables of session one participants. During the quiet condition both participants conversed in quiet and in the symmetrical condition both conversed in noise, while during the asymmetrical condition only the right booth participant was subjected to noise and the left booth participant conversed without any background noise.



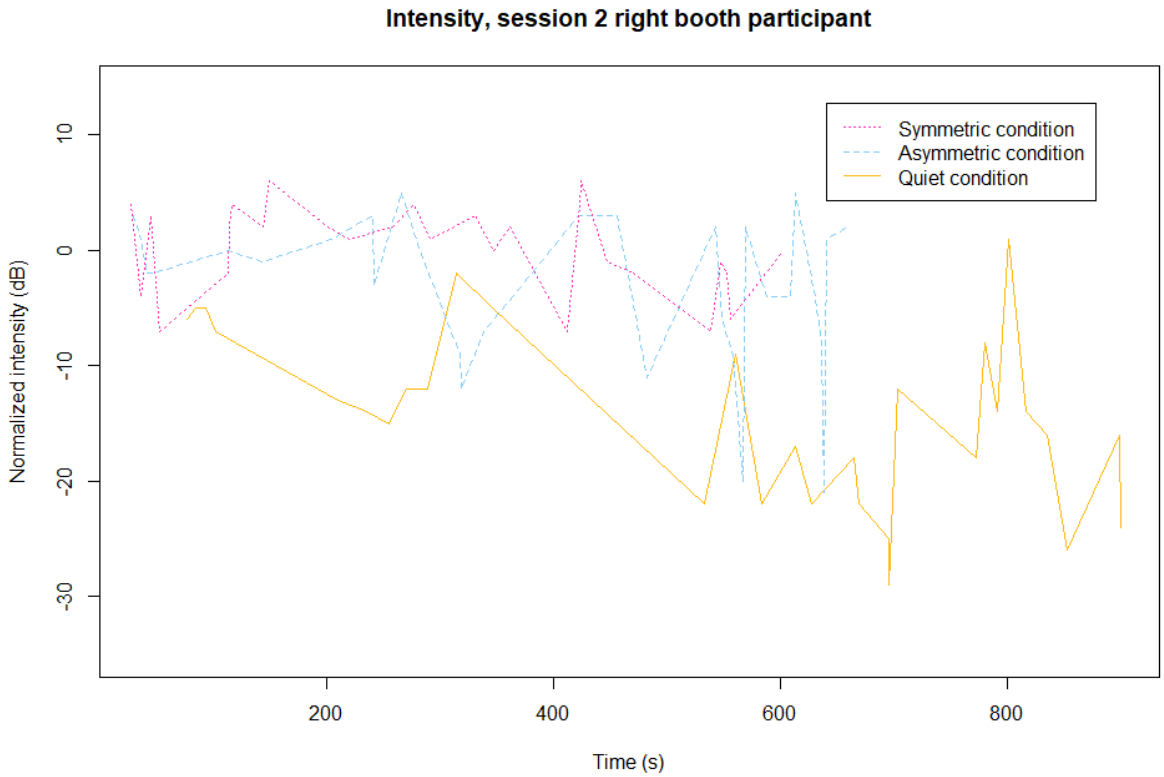


Figure 3. Intensity data time series of target syllables of the participants in session two. During the quiet condition both participants conversed in quiet and in the symmetrical condition both conversed in noise, while during the asymmetrical condition only the right booth participant was subjected to noise and the left booth participant conversed without any background noise.

minimal convergence of range between the quiet condition and the other conditions. Especially for the left booth participant of session one (Figure 2) who was not subjected to noise during the asymmetrical condition, the values of the asymmetrical condition seem to be situated mainly between the values of the quiet and the symmetrical condition but overlapping more with the quiet condition. For the other participants, the data of the symmetrical and the asymmetrical condition appear to be concentrated on similar areas with each other, but with the symmetrical condition data sitting higher for the most part. The largest oscillations in the data across speakers appear to happen during the quiet condition while data of the symmetrical condition is more stationary, which can be expected from Lombard speech.

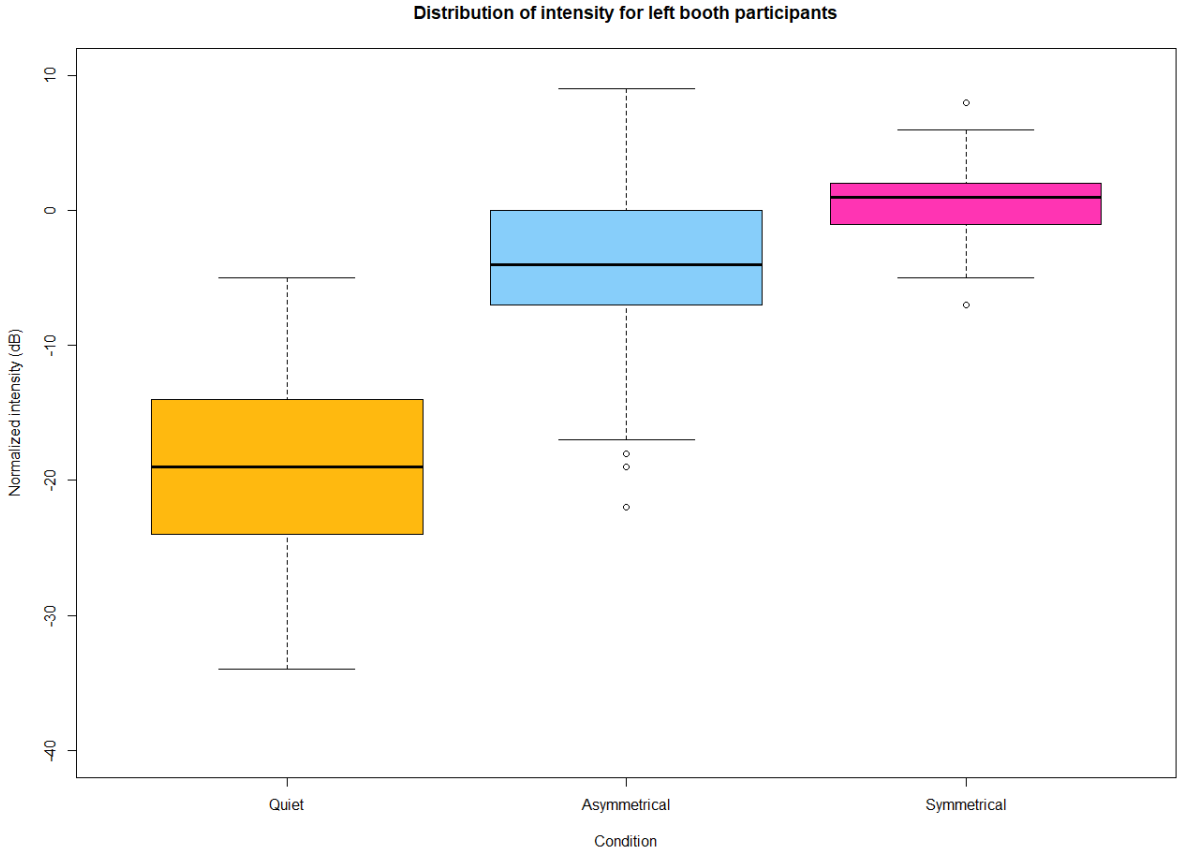


Figure 4. Distribution of intensity data gathered in (1) quiet, in (2) the asymmetrical condition where the participant in question was conversing in silence while the other interlocutor was subjected to noise and in (3) the symmetrical condition where both interlocutors were subjected to noise. The normalization of values was carried out by calculating the difference between the mean intensity level of an individual speaker in the symmetrical condition and the value in a compared condition. Left booth participant refers to the role of the participant in the asymmetrical condition.

4.3.2 Intensity variations of groups

Left booth participants			
	Quiet	Asym	Sym
Quiet	xxxxxx	p < 0.0001	p < 0.0001
Asym	p = 0.162	xxxxxx	p < 0.0001
Sym	p < 0.0001	p < 0.0001	xxxxxx

Right booth participants			
	Quiet	Asym	Sym
Quiet	xxxxxx	p < 0.0001	p < 0.0001
Asym	p = 0.033	xxxxxx	p = 0.004
Sym	p < 0.0001	p = 0.036	xxxxxx

Table 2. Welch's two-sample t-test and F-test p-values between the intensity data of different conditions of the left and right booth participants. The F-test values are presented on the white background and the t-test values on the grey background. "Quiet" refers to the quiet condition where both participants of a pair conversed in quiet, "Sym" refers to the symmetrical condition where both interlocutors were conversing in noise and "Asym" refers to the asymmetrical condition where the right booth participants were subjected to background noise while the left booth participants conversed in silence.

Based on the differences observed in the time series data visualizations, the means and variances of the different conditions were compared. To statistically test the differences in means, Welch's two-sample t-tests were utilized and F-tests were carried out to compare variances. The distribution of the intensity data from participants who during the asymmetrical condition were in silence (left booth participants) can be found in Figure 4 and the p-value results of t-tests and F-tests between the three different conditions can be found in Table 2. The estimated mean intensity of the speech produced in the quiet condition by the left booth participants is at -19 dB, which is the largest contrast from the symmetrical condition. The estimated mean intensity in the asymmetrical condition is at -5 dB, which is somewhat closer to the symmetrical than the quiet condition. The symmetrical condition, being the point of reference, is near 0 dB. Welch two-sample t-tests show that the differences between all the three conditions are all statistically significant with $p < 0.0001$. The data from the symmetrical condition shows the least amount of variation in intensity, which was already implied in the temporal data. Both the asymmetrical and the quiet condition show similar wideness of range in the intensity data and an F-test reveals the variances of the two conditions are not significantly different ($p = 0.162$). The variation of the quiet condition data could however be due to the inter-speaker differences contained in the data instead of the individual quiet condition data being varied.

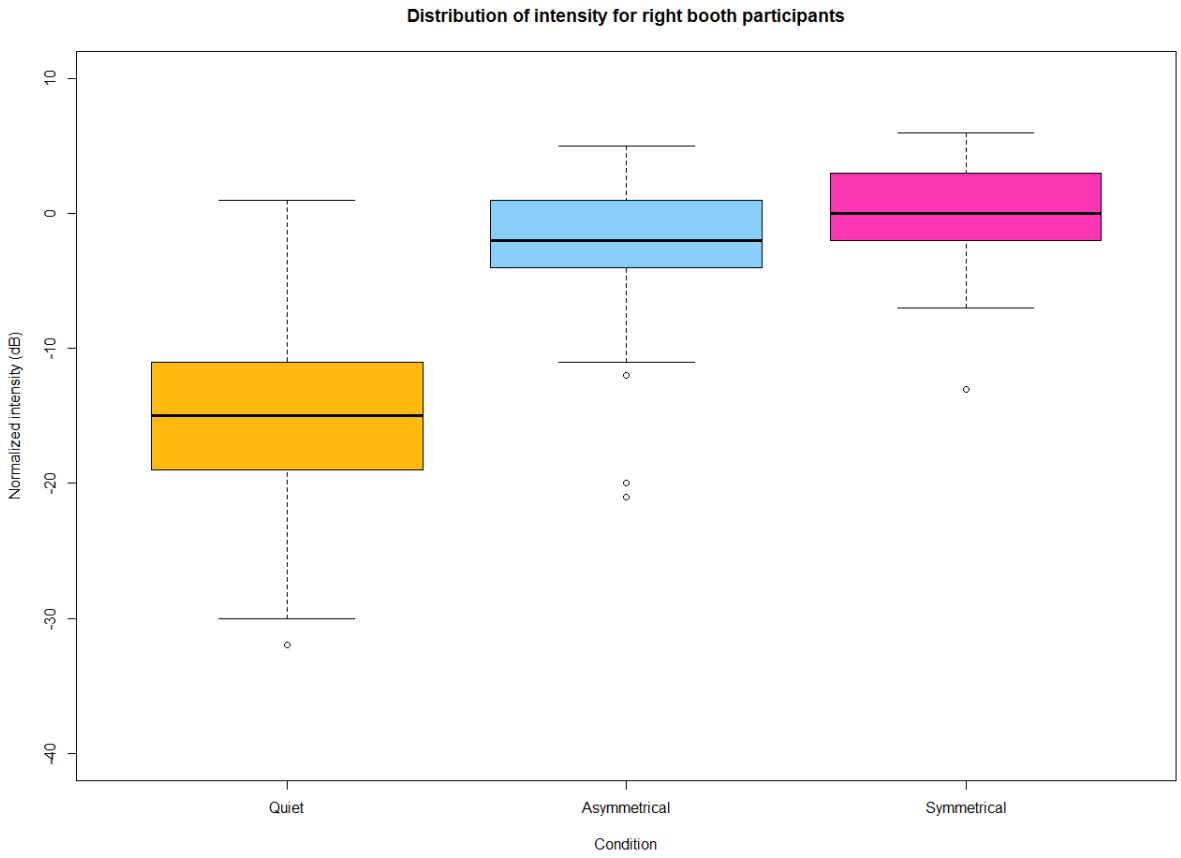


Figure 5. Distribution of intensity data gathered in (1) quiet, in (2) the asymmetrical condition where the participant in question was conversing in noise while the other interlocutor was in quiet and in (3) the symmetrical condition where both interlocutors were subjected to noise. The normalization of values was carried out by calculating the difference between the mean intensity level of an individual speaker in the symmetrical condition and the value in a compared condition. Right booth participant refers to the role of the participant in the asymmetrical condition.

Although showing a similar kind of trend, the intensity data gathered from right booth participants who during the asymmetrical condition were subjected to noise (Figure 5) differs somewhat from the data that was gathered from the left booth participants who were not in noise during the asymmetrical condition. The estimated mean intensity of the right booth participants in the quiet condition is at -15 dB. According to Welch's t-tests (Table 2), the difference between the quiet and the asymmetrical condition of the right booth participants is expectedly statistically significant with a p-value < 0.0001. The mean intensity in the asymmetrical condition of the right booth participants is only -3 dB from the symmetrical condition which once again is at around 0 dB. Seen as the right booth participants were

producing Lombard speech in both the asymmetrical and the symmetrical condition, it could be expected that the difference in the means of the two conditions would not be statistically significant. However, the t-test between the conditions shows the difference to be statistically significant although not as strongly as between the other conditions, with p = 0.004.

4.3.3 Individual variations of intensity

Session 1 left booth participant			
	Quiet	Asym	Sym
Quiet	xxxxxx	p < 0.0001	p < 0.0001
Asym	p = 0.136	xxxxxx	p < 0.0001
Sym	p = 0.0002	p = 0.01	xxxxxx

Session 1 right booth participant			
	Quiet	Asym	Sym
Quiet	xxxxxx	p < 0.0001	p < 0.0001
Asym	p = 0.0003	xxxxxx	p = 0.07
Sym	p = 0.0164	p = 0.265	xxxxxx

Session 2 left booth participant			
	Quiet	Asym	Sym
Quiet	xxxxxx	p < 0.0001	p < 0.0001
Asym	p = 0.03	xxxxxx	p = 0.224
Sym	p < 0.0001	p = 0.027	xxxxxx

Session 2 right booth participant			
	Quiet	Asym	Sym
Quiet	xxxxxx	p < 0.0001	p < 0.0001
Asym	p = 0.43	xxxxxx	p = 0.024
Sym	p = 0.0007	p = 0.007	xxxxxx

Table 3. Welch's two-sample t-test and F-test p-values between the intensity data of different conditions of the individual participants. The F-test values are presented on the white background and the t-test values on the grey background. "Quiet" refers to the quiet condition where both participants of a pair conversed in quiet, "Sym" refers to the symmetrical condition where both interlocutors were conversing in noise and "Asym" refers to the asymmetrical condition where the right booth participants were subjected to background noise while the left booth participants conversed in silence.

Some differences between the participants can be observed in the individual speaker data of intensity (Figure 6). The F-test and t-test results of the individual intensity data can be found in Table 3. The quiet condition shows the widest range of values across speakers as expected, confirming that the variance observed in the quiet condition data of the left booth participants is indeed due to variation in the individual data. Looking at the intensity data of the quiet condition in session two, the right booth participant has used her voice in a more varied manner compared to the interlocutor in the left booth whose values are more closely centered around the -20 dB mark. The participant did raise her voice from the quiet condition,

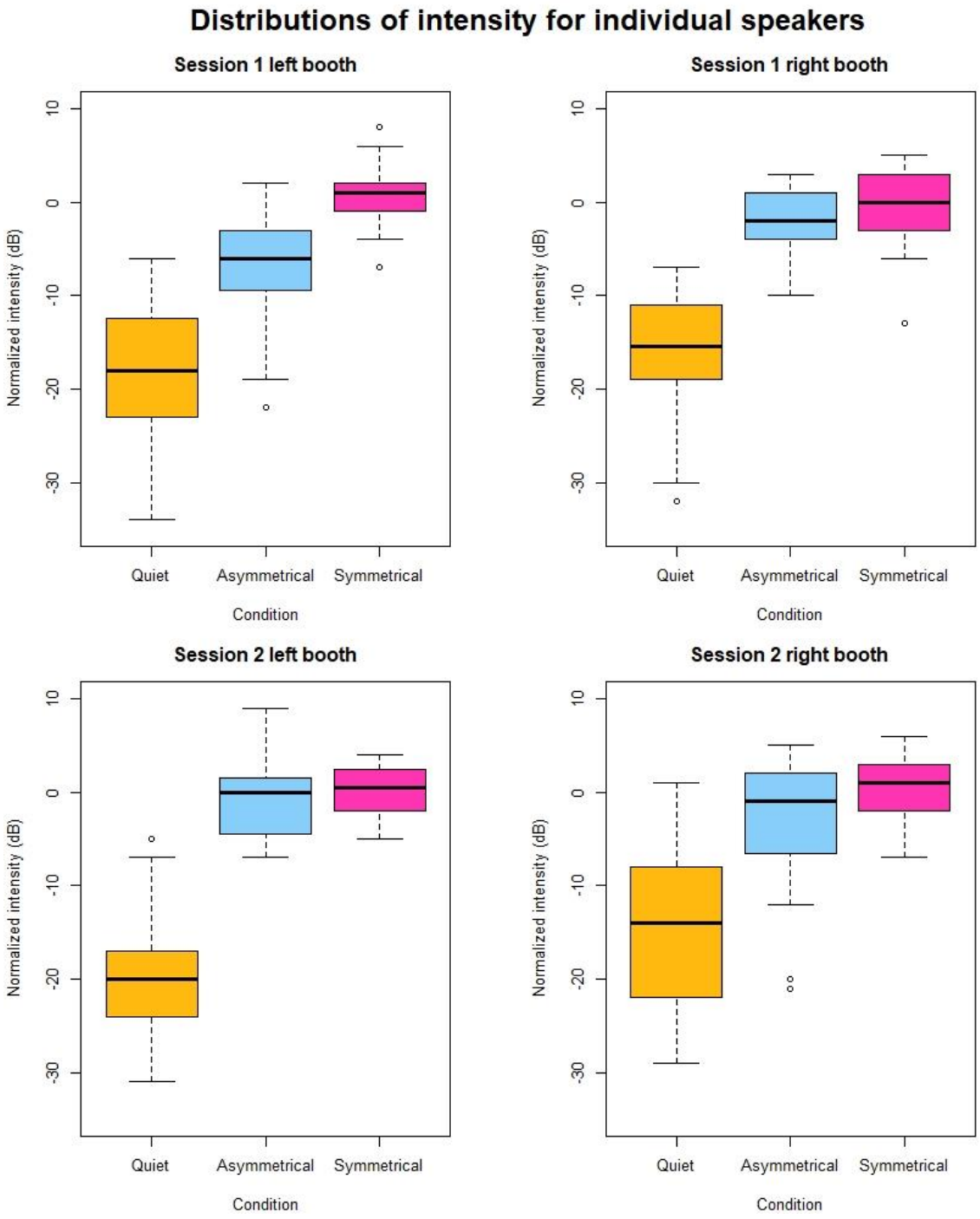


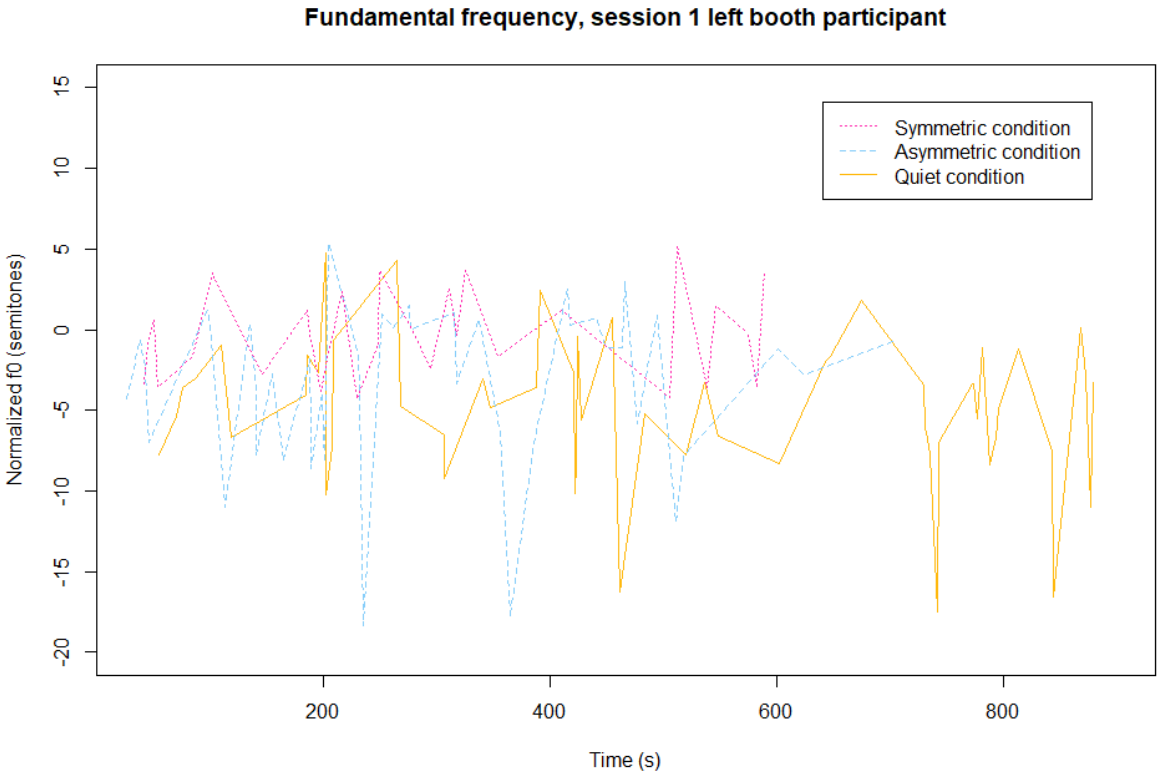
Figure 6. Intensity distributions of individual speakers. Each session had a pair of speakers, one in the left booth and one in the right (indicated in the plots). The participants conversed in three different background conditions: (1) in quiet, (2) in the asymmetrical condition where the left booth participants were in silence and the right booth participants were subjected to noise and (3) in the symmetrical condition where both participants were in noise. The normalization of values was carried out by calculating the difference between each participant's mean intensity level in the symmetrical condition and the intensity level of a given condition.

but not nearly to the levels that she utilized in the Lombard speech of the symmetrical condition. This observation is in line with the participant reporting that she occasionally forgot that the other interlocutor was subjected to noise and therefore did not speak as loudly as she could have. The left booth participant of session two however shows no statistically significant difference between the intensity of the symmetrical and the asymmetrical condition ($p = 0.224$). In fact, the intensity of the asymmetrical condition of this participant ranges even higher than the mean intensity level of the symmetrical condition, although the values do concentrate around the 0 dB mark and show that the participant held a raised voice a bit more consistently compared to the other left booth participant. Both going under the intensity levels that were utilized during the symmetrical condition and going over them are understandable responses to attempting to "calibrate" one's vocal production to a level that would suit the new unbalanced situation at hand.

4.3.4 Temporal variations of fundamental frequency

Figures 7 and 8 present the changes in fundamental frequency in time for individual speakers. Much like with the intensity data, there are broad oscillations in the fundamental frequency data due to the data expanding a long window of time and presenting only the target syllables. Once again, no clear temporal trends occur in the data to make note of, such as overall increases or decreases in f_0 over time.

For most speakers, the f_0 data appears to be situated more closely together across the three conditions compared to the intensity data which showed distinct contrasts between the conditions. Especially the two right booth participants who during the asymmetrical condition were subjected to noise show data whose ranges are fairly overlapping. For example, the data of all three conditions from the right booth participant of session one (Figure 7) populates the area between -5 and 5 semitones rather densely. However, the f_0 data of the left booth participants who during the asymmetrical condition were not subjected to noise appear to have somewhat clearer distinctions between conditions.



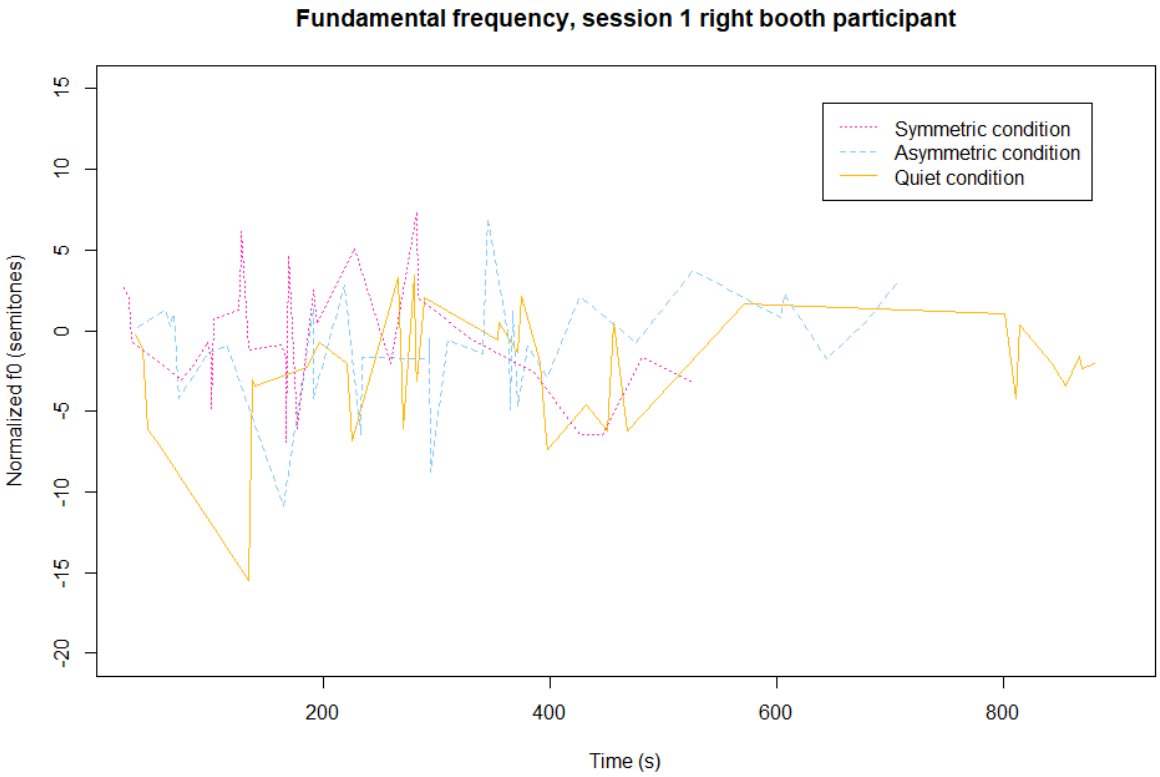
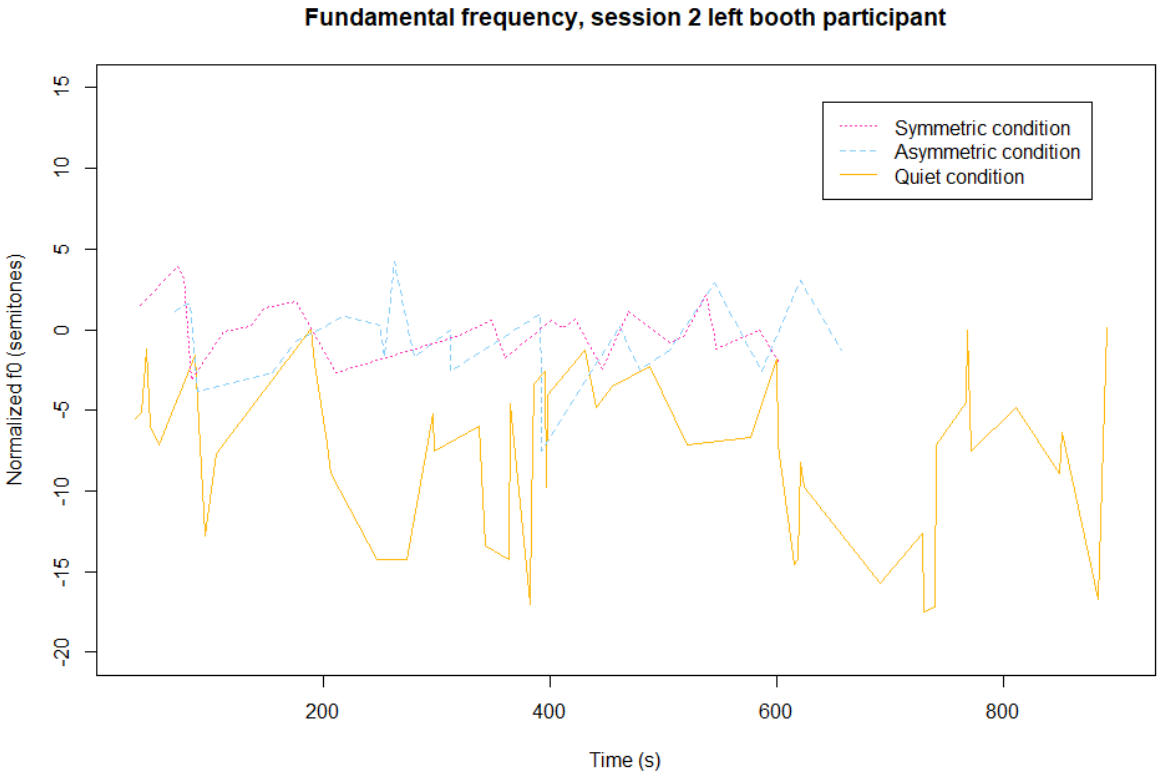


Figure 7. The fundamental frequency data time series of target syllables in session one. During the quiet condition both participants conversed in quiet and in the symmetrical condition both conversed in noise, while during the asymmetrical condition only the right booth participant was subjected to noise and the left booth participant conversed without any background noise.



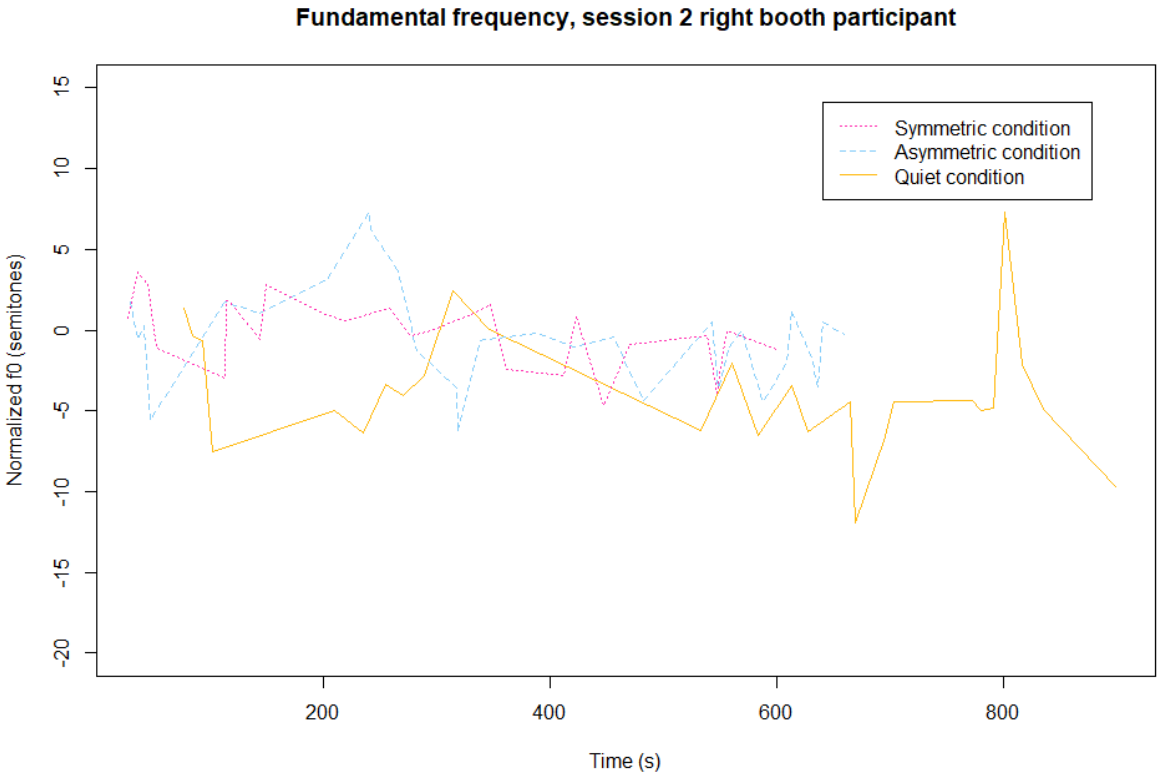


Figure 8. The fundamental frequency data time series of target syllables in session two. During the quiet condition both participants conversed in quiet and in the symmetrical condition both conversed in noise, while during the asymmetrical condition only the right booth participant was subjected to noise and the left booth participant conversed without any background noise.

The f_0 data of the quiet condition of the left booth participant of session two (Figure 8) is visibly situated lower compared to the data of the symmetric and the asymmetric conditions, which appear to share a similar range with each other. The data of the left booth participant of session one (Figure 7) on the other hand shows more similarity of range between the asymmetrical and the quiet condition, while the data of the symmetrical condition is situated higher compared to the other two conditions.

4.3.5 Fundamental frequency variations of groups

Left booth participants			
	Quiet	Asym	Sym
Quiet	xxxxxx	p < 0.0001	p < 0.0001
Asym	p = 0.391	xxxxxx	p = 0.0005
Sym	p < 0.0001	p < 0.0001	xxxxxx

Right booth participants			
	Quiet	Asym	Sym
Quiet	xxxxxx	p = 0.0003	p < 0.0001
Asym	p = 0.509	xxxxxx	p = 0.59
Sym	p = 0.111	p = 0.342	xxxxxx

Table 4. Welch's two-sample t-test and F-test p-values between the fundamental frequency data of different conditions of the left and right booth participants. The F-test values are presented on the white background and the t-test values on the grey background. "Quiet" refers to the quiet condition where both participants of a pair conversed in quiet, "Sym" refers to the symmetrical condition where both interlocutors were conversing in noise and "Asym" refers to the asymmetrical condition where the right booth participants were subjected to background noise while the left booth participants conversed in silence.

Distributions of the fundamental frequency data of the left booth participants who were in quiet during the asymmetrical condition follow a similar pattern to the corresponding intensity data (Figure 9, Table 4). The estimated mean of the quiet condition for these participants is -6.2 semitones, and the estimated mean of the asymmetrical condition is -2.4 semitones. The difference is statistically significant with $p < 0.0001$. The mean of the symmetrical condition is at -0.1 semitones, leaving only a difference of 2.3 semitones between the asymmetrical and the symmetrical condition, which however is statistically significant with a value of $p = 0.0005$. Once again, the quiet condition has allowed for the greatest variation in the fundamental

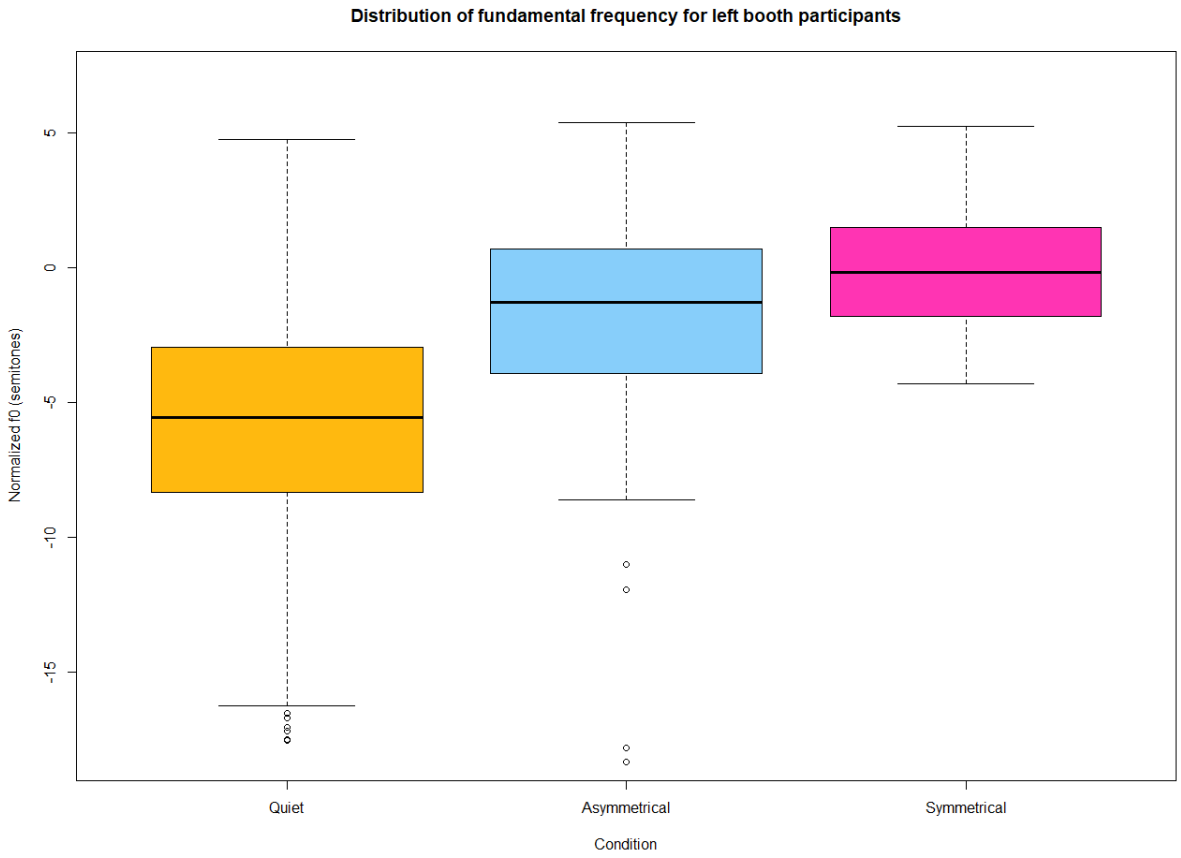


Figure 9. The distribution of f_0 *data in (1) quiet, in (2) the asymmetrical condition where the participant in question was conversing in silence while the other interlocutor was subjected to noise and in (3) the symmetrical condition where both interlocutors were subjected to noise. Left booth participant refers to the position that the participant was in during the asymmetrical condition. The normalization of values was carried out by calculating the difference between the mean* f_0 *of the symmetrical condition of the individual speaker and the* f_0 *of a given condition.*

frequencies, and the Lombard speech of the symmetrical condition has led to more closely distributed data which often is the case with such speech that is produced in the restrictions of a loud environment.

In the case of fundamental frequency, the data of the right booth participants who were subjected to noise during the asymmetrical condition is more evenly distributed across the three conditions (Figure 10, Table 4). The estimated mean of f_0 in the quiet condition is -3.0 semitones. The estimated mean of the asymmetrical condition is only -0.6 semitones whereas the estimated mean of the symmetrical condition is -0.3 semitones. The difference between the quiet and the asymmetrical condition is statistically significant at a level of $p = 0.0003$ while the difference between the asymmetrical and the symmetrical condition is unsurprisingly not significant at $p = 0.59$. The f_0 data of the asymmetrical condition is seemingly more concise

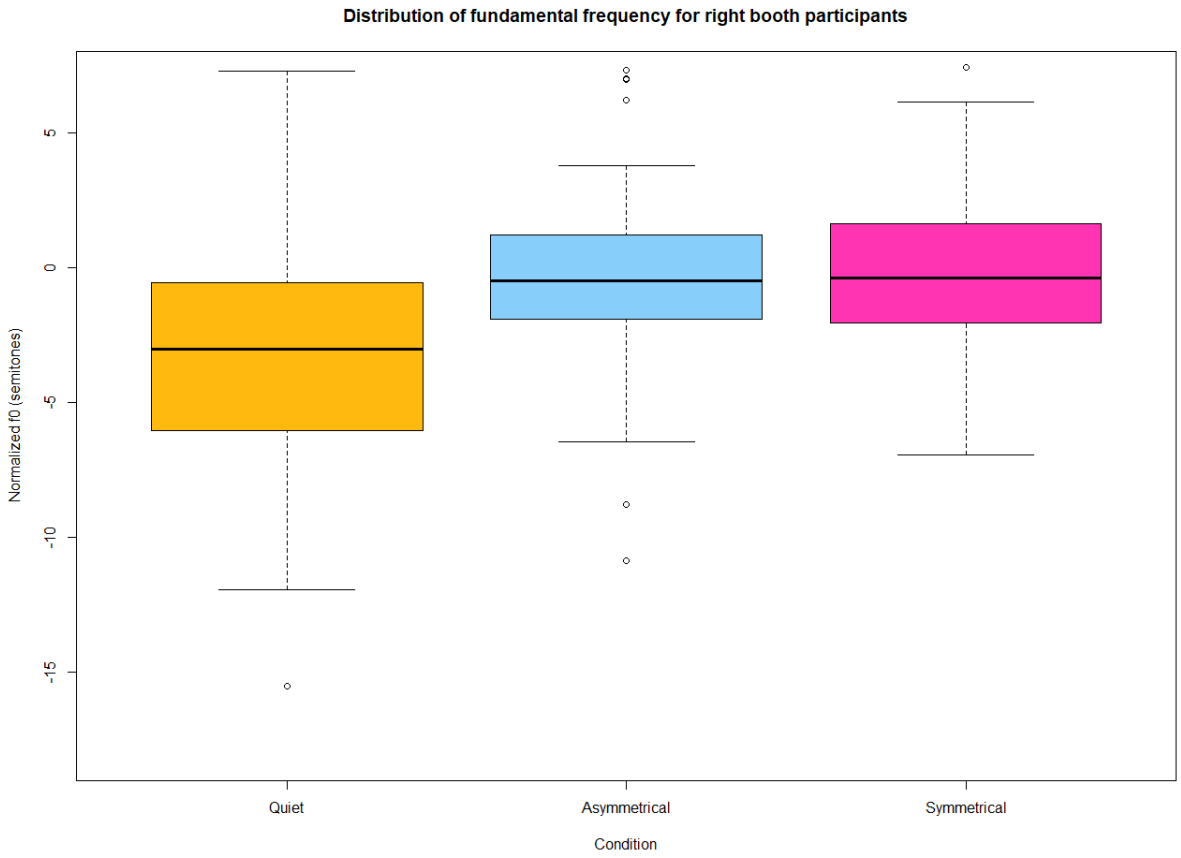


Figure 10. The distribution of f_0 *data in (1) quiet, in (2) the asymmetrical condition where the participant in question was conversing in noise while the other interlocutor was in quiet and in (3) the symmetrical condition where both interlocutors were subjected to noise. Right booth participant refers to the position that the participant was in during the asymmetrical condition. The normalization of values was carried out by calculating the difference between the mean* f_0 *of the symmetrical condition of the individual speaker and the* f_0 *of a given condition.*

than that of the symmetrical condition, though it does contain a number of outliers. The variances of the symmetrical and the asymmetrical condition are not significantly different ($p = 0.342$).

4.3.6 Individual variations of fundamental frequency

The f_0 data of individual speakers (Figure 11) goes mainly hand in hand with the individual intensity data. The right booth participant of session one shows the most consistent f_0 data out of all the participants, with no statistically significant differences between the symmetrical and the asymmetrical condition or between the quiet and the asymmetrical condition, and only a slightly significant difference between the quiet and the symmetrical condition (Table 5). The

Session 1 left booth participant			
	Quiet	Asym	Sym
Quiet	xxxxxx	p = 0.079	p < 0.0001
Asym	p = 0.454	xxxxxx	p = 0.001
Sym	p = 0.008	p = 0.002	xxxxxx

Session 1 right booth participant			
	Quiet	Asym	Sym
Quiet	xxxxxx	p = 0.05	p = 0.03
Asym	p = 0.871	xxxxxx	p = 0.777
Sym	p = 0.853	p = 0.978	xxxxxx

Session 2 left booth participant			
	Quiet	Asym	Sym
Quiet	xxxxxx	p < 0.0001	p < 0.0001
Asym	p = 0.001	xxxxxx	p = 0.272
Sym	p < 0.0001	p = 0.066	xxxxxx

Session 2 right booth participant			
	Quiet	Asym	Sym
Quiet	xxxxxx	p = 0.001	p < 0.0001
Asym	p = 0.294	xxxxxx	p = 0.597
Sym	p = 0.002	p = 0.029	xxxxxx

Table 5. Welch's two-sample t-test and F-test p-values between the f_0 *data of different conditions of the individual participants. The F-test values are presented on the white background and the t-test values on the grey background. "Quiet" refers to the quiet condition where both participants of a pair conversed in quiet, "Sym" refers to the symmetrical condition where both interlocutors were conversing in noise and "Asym" refers to the asymmetrical condition where the right booth participants were subjected to background noise while the left booth participants conversed in silence.*

estimated mean f_0 of this participant in the asymmetrical condition is -0.4 semitones which is a bit higher than in the symmetrical condition, which is at -0.7 semitones. This is the only instance of the estimated mean of the asymmetrical condition situating higher than the estimated mean in the symmetrical condition. Perhaps the most interesting deviation in the data of this participant is the lack of any greater increase in fundamental frequency from speaking in silence to speaking in noise. Classically overall increases in f_0 are connected to the production of Lombard speech, which the participant did in both the asymmetrical and the symmetrical condition, yet the f_0 values of this particular participant stay similar across the three conditions, with no statistically significant differences between their variances. The right booth participant of session two has in turn talked with an overall higher pitch in the asymmetrical and the symmetrical condition (both in which she produced Lombard speech) compared to the quiet condition, which is an expected outcome when comparing normal and Lombard speech.

The left booth participants of session one and two who were in quiet during the asymmetrical condition show some differences between each other as well. The f_0 data of the left booth participant of session one varies from 5 semitones to -10 semitones with a few outliers situated under -15 semitones in the quiet condition. The corresponding data of the left booth participant in session two ranges down to -18 semitones, which is 1.5 octaves lower than the baseline of 0 semitones, the maximum for this speaker in the quiet condition.

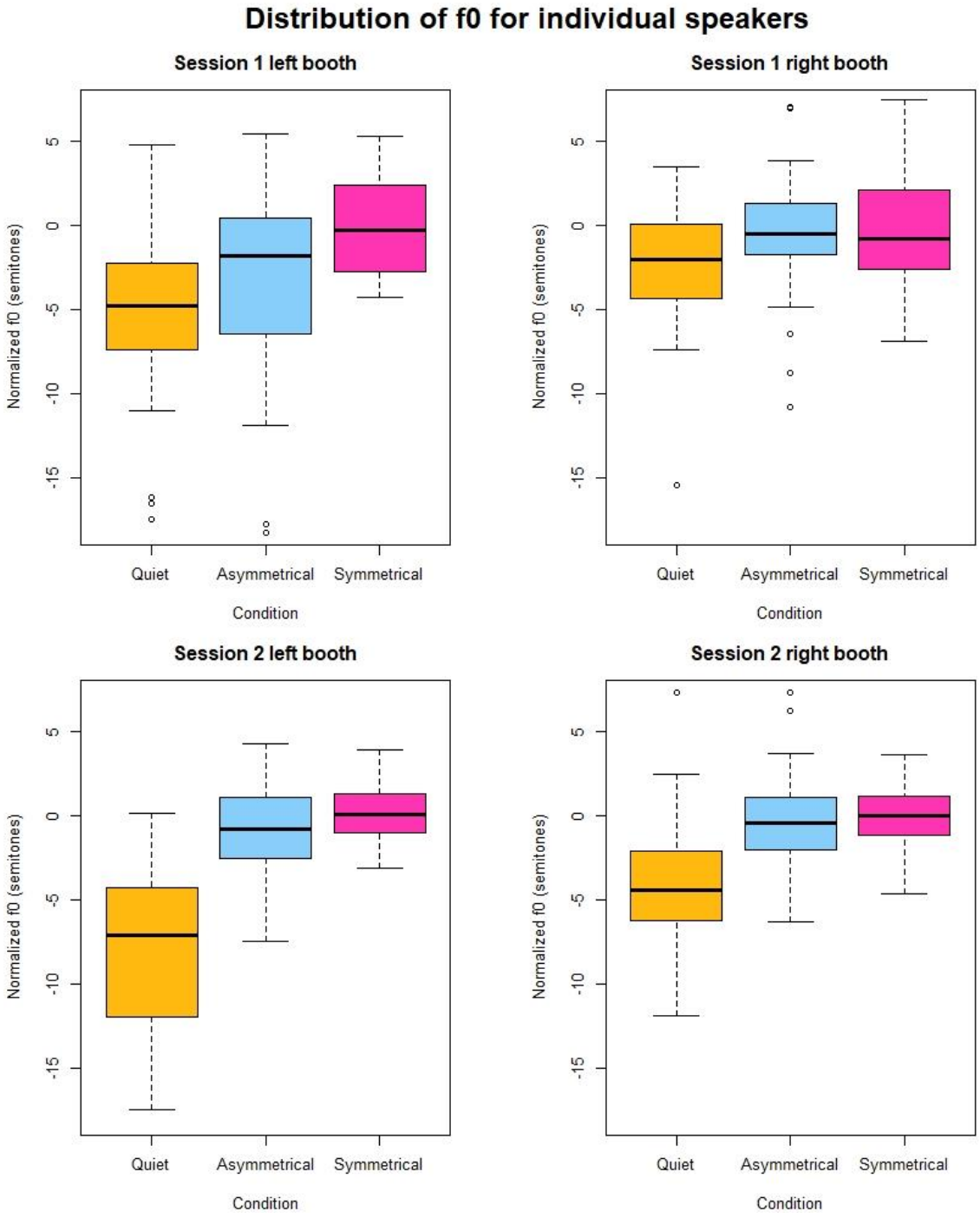


Figure 11. Fundamental frequency distributions of individual speakers. Each session had a pair of speakers, one in the left booth and one in the right (indicated in the plots). The participants conversed in three different background conditions: (1) in quiet, (2) in the asymmetrical condition where the left booth participants were in silence and the right booth participants were subjected to noise and (3) in the symmetrical condition where both participants were in noise. The normalization of values was carried out by calculating the difference between each participant's mean f_0 *in the symmetrical condition and the* f_0 *of a given condition.*

Seen as the ranges of f_0 values of the left booth speakers are situated on different areas in the quiet condition, it is no surprise that there are differences in how they compare to the corresponding data in the asymmetrical condition. There is no statistically significant difference between the quiet and the asymmetrical condition data of the left booth participant of session one ($p = 0.08$), and the asymmetrical data of the participant ranges even lower than the data of the quiet condition. An F-test comparing the two conditions also shows no statistically significant difference between variances of the asymmetrical and the quiet condition ($p = 0.454$). This is a notable difference when compared to the left booth participant of session two, who shows a clear distinction between the quiet and the asymmetrical condition, which was already visible in the temporal visualization of the f_0 data of this speaker. A Welch two-sample t-test shows the difference between the quiet and the asymmetrical condition data of this speaker to be statistically significant with $p < 0.0001$.

Whereas the left booth participant of session two has no statistically significant difference between the f_0 data of the asymmetrical and the symmetrical condition ($p = 0.272$), the left booth participant in session one shows a significant difference between the symmetrical and the asymmetrical condition ($p = 0.001$). The f_0 data of the asymmetrical and the symmetrical conditions of the left booth participant of session two are clearly situated together further away from the quiet condition, much like in the case of intensity, while the asymmetrical f_0 data of the left booth participant of session one sits more evenly between the quiet and the symmetrical condition.

4.4 Discussion

As was expected, differences in the acoustics of speech and the intelligibility of speech were observed when either both interlocutors were in noise, or when just one of the interlocutors was subjected to noise. Though the left booth participants who were not in noise during the asymmetrical condition did on average talk louder in the asymmetrical condition than in the quiet condition, the speech produced in the symmetrical condition was on average the loudest. In the individual intensity data however the difference between the asymmetrical and the symmetrical data was significant for only one of the left booth participants, and for the other left booth participant there was no significant difference between the symmetrical and the asymmetrical conditions. The fundamental frequency data of these participants follows the

intensity data closely with the quiet condition yielding on average the lowest values, the asymmetrical condition the second highest and the symmetrical condition the highest f_0 values.

A similar trend to the intensity data of the left booth participants can be seen in the data of the right booth participants who were subjected to noise during the asymmetrical condition. These participants likewise demonstrated an expected difference between the intensity data of the normal speech of the quiet condition and the Lombard speech of the symmetrical condition, but unexpectedly also produced on average lower intensity values in the asymmetrical condition than in the symmetrical condition. For the right booth participants, the f_0 values of the three conditions are more closely together with no significant differences between the asymmetrical and the symmetrical condition.

In the quiet condition the absence of background noise allowed for the use of very quiet vocalizations in addition to using louder vocalizations at times, contributing to the larger ranges in intensity in the quiet condition. Larger ranges of f_0 were also recorded in the quiet condition, which likely ties in with the changes in intensity. The background noise of the symmetrical condition created the opposite effect compared to the quiet condition, with values of intensity grouping closer together as the result of Lombard speech.

The left booth participants, that is, the participants that were in quiet trying to communicate to their partner who was in noise during the asymmetrical condition, sometimes forgot to increase their vocal efforts to a sufficient level when their interlocutor was in noise and the left booth participant of session one even reported occasionally forgetting about the background noise that the other interlocutor was subjected to. The participants in the right booth who were subjected to noise in both the symmetrical and the asymmetrical condition in turn reported their interlocutors as being hard to hear in the asymmetrical condition. The participants in the left booth however did increase their vocal efforts at times where they thought to have found solutions for the sudoku. This can be seen in the greater variation of f_0 and intensity data of the left booth participants in the asymmetrical condition compared to the symmetrical condition. Contradictory to what was predicted, target syllables were misheard by the right booth participants (or by any participants at that) only on one occasion, during the symmetrical condition in session one when both interlocutors were subjected to noise. This would indicate that at least at the background noise level that was present in this study, during the asymmetrical condition the left booth participants were able to make ample changes to their vocal production

without major disturbances in the flow of the conversation when it came to conveying the right target syllables.

Right booth participants who were subjected to noise during the asymmetrical condition showed clear distinction between the quiet and the asymmetrical condition intensity data, but also distinctions between the intensities of the symmetrical and the asymmetrical condition. Although producing Lombard speech in both conditions, these participants likely did not speak with as high intensity in the asymmetrical condition due to some of the pressures of producing intelligible speech taken away by the fact that the person in quiet would likely hear them no matter what, yet still could not ignore the effects of the background noise. This demonstrates just how strong the auditory feedback loop is and how deeply the Lombard effect affects our speech production, and in part discredits the notion by Lane & Tranel (1971) of the Lombard effect only having roots in the conscious efforts for intelligibility in one's speech production.

The fact that in the asymmetrical condition both the interlocutor in noise and the interlocutor in silence started producing speech at an intensity level closer to their partner might also have to do with the phenomenon of entrainment, in which interlocutors engaged in a conversation start to behave more like their conversational partner (Gregory et al., 1993; Levitan et al., 2016). This also extends to speech and has been recorded to affect for example the intensity level and the f_0 of speakers (Levitan & Hirschberg, 2011; Natale, 1975).

As noted earlier, the variations in f_0 that happen due to the Lombard effect are in part connected to the nature of phonation during Lombard speech in which the increased subglottal pressure also increases f_0. In addition, the presence of noise can lead to speakers emphasizing their speech more thus resulting in heightened f_0. Across the data, f_0 values showed similar trends to the intensity values, which is in line with previous findings (e.g., Lu & Cooke, 2009a). The left booth participant of session two showed greater difference between the f_0 data of the quiet condition and the symmetrical and asymmetrical conditions compared to other participants (Figure 11). This distinction could be brought on by a difference in background: the other three participants had a background in singing which would explain their better control over not increasing their fundamental frequency as a result of increased intensity.

While it was speculated that the people in silence trying to converse with a person in noise could resort to shouting when the Lombard effect was not present and they could only really rely on the public auditory feedback loop, the conditions of communicating were not harsh

enough for the people in silence to resort to such a technique, i.e., the background noise was not so overwhelming that it would have drowned out quieter speech entirely. If the differences between Lombard speech and shouted speech want to be studied further in the future, harsher conditions need to be implemented to elicit the shouting in the asymmetrical condition. The easiest way to do this would be to increase the sound pressure level of the background noise, in which case more information would be lost if the people conversing in silence spoke at all too quietly. In addition to altering the speech conditions, some fluctuations in the target syllable frequencies could possibly be managed through designing the sudoku tasks used in this kind of an experimental setting more carefully. By making sure the sudokus have equal distributions of missing syllables, the resulting data could be more equally distributed, assuming that there are correlations between the syllables that need to be filled into a sudoku and the syllables the participants converse about. This sort of deeper analysis of the research method is however outside the scope of this paper.

The right booth participant of session two who was subjected to noise in the asymmetrical condition reported trying to convey visual cues to her interlocutor through the booth window, but the partner in turn had not made any notice of this. Altogether the participants of both sessions were largely engaging in the conversation with their gaze focused on the sudokus and only briefly glancing at their interlocutor from time to time. It has been demonstrated that the absence or presence of a visual connection between interlocutors can affect both auditory and visual Lombard speech production (Fitzpatrick et al., 2015). The merging of the two topics, the asymmetrical Lombard effect and the variance in visual connection, could therefore provide a possible subject for future research.

5 Conclusions

Compared to the Lombard speech of the symmetrical condition and the speech in silence in the quiet condition, the unbalanced speaking conditions that the participants were subjected to in the asymmetrical condition clearly evoked changes in the speech production of both the people who were in noise and the people who conversed in silence. Although the left booth participants spoke in silence during the asymmetrical condition, they increased their vocal effort for the sake of their interlocutor, and the right booth participants, although under the Lombard effect, spoke more quietly when talking in noise to a person who was not in noise. No second-hand

Lombard effect was discovered, that is, that the left booth participants would have consistently increased their vocal efforts due to the right booth participants raising their voice. In fact, vocal efforts were increased by the left booth participants usually only in places where the information communicated was regarded as crucial to the solving of the task or when the interlocutor in noise requested their partner to speak louder.

The experimental setting used in this study shows great promise for collecting spontaneous conversational speech data for other kinds of experiments as well. Due to the speech collected for this experiment being conversational, it contains interesting traits of dialogical speech like turn taking, entrainment in conversational partners and even laughter that could be researched further.

What this study demonstrated was that once the conditions of communication were made more challenging, neither of the interlocutors in a pair continued to produce exactly the kind of speech that their respective sound environments required but rather readjusted their speech production to a level that was both intelligible for the other participant and maintainable for a longer period of time. More importantly, it showed that not only does the communicative aspect of a speech situation increase the effects of the Lombard effect, but it can also diminish them. To further research this phenomenon and to make up for individual differences between speakers, further analysis of the data will be carried out and more data will be gathered in the future.

References

Boersma, P. & Weenik, D. (2019). Praat (version 6.1.02) (computer program), from http://www.fon.hum.uva.nl/praat/.

Bottalico, P., Passione, I.I., Graetzer, S. & Hunter, E.J. (2017). Evaluation of the starting point of the Lombard effect. *Acta Acoustics united with Acustica, 103*(1), 169 –172. https://doi.org/10.3813/AAA.919043

Brumm, H. & Zollinger, S.A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour, 148*(11–13), 1173–1198. https://doi.org/10.1163/000579511X605759

Charlip, W.S. & Burke, K.W. (1969). Effects of noise on selected speech parameters. *Journal of Communication Disorders, 2*(3), 212–219. https://doi.org/10.1016/0021-9924(69)90016-1

Cooke, M., Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America, 128,* 2059–2069. https://doi.org/10.1121/1.3478775

Cvejic, E., Kim, J., & Davis, C., (2012). Effects of seeing the interlocutor on the production of prosodic contrasts (L). *The Journal of the Acoustical Society of America, 131*(2), 1011–1014. https://doi.org/10.1121/1.3676605

Dawson, C., Aalto.d D., Šimko, J., Vainio, M. (2017). The influence of fundamental frequency on perceived duration in spectrally comparable sounds. PeerJ 5:e3734. https://doi.org/10.7717/peerj.3734

Farley, S.D., Hughes, S.M., LaFayette, J.N. (2013). People Will Know We Are in Love: Evidence of Differences Between Vocal Samples Directed Toward Lovers and Friends. *Journal of Nonverbal Behavior, 37*, 123–138. https://doi.org/10.1007/s10919-013-0151-3

Fitzpatrick, M., Kim, J., & Davis, C. (2015). The effect of seeing the interlocutor on auditory and visual speech production in noise. *Speech Communication, 74*, 37–51. https://doi.org/10.1016/j.specom.2015.08.001

Garnier, M., Henrich, N. & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language and Hearing Research, 53*(3), 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138)

Garnier, M. & Henrich, N. (2013). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language, 28*(2), 580–597. https://doi.org/10.1016/j.csl.2013.07.005

Garnier, M., Ménard, L., & Alexandre, B. (2018). Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues? *The Journal of the Acoustical Society of America*, *144*(2), 1059–1074. https://doi.org/10.1121/1.5051321

Gregory, S., Webster, S., & Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language & Communication*, *13*(3), 195–217. https://doi.org/10.1016/0271-5309(93)90026-J

Huber, J.E., Chandrasekaran, B., & Wolstencroft, J.J. (2005). Changes to respiratory mechanisms during speech as a result of different cues to increase loudness. *Journal of Applied Physiology, 98*, 2177–2184. https://doi.org/10.1152/japplphysiol.01239.2004

Junqua, J.C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, *93*, 510–524. https://doi.org/10.1121/1.405631

Lane, H. L., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research, 14*, 677–709. https://doi.org/10.1044/jshr.1404.677

Leonard, M. L., & Horn, A. G. (2005). Ambient noise and the design of begging signals. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1563), 651–656. https://doi.org/10.1098/rspb.2004.3021

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Interspeech 2011*, 3081–3084. Florence, Italy. https://doi.org/10.21437/Interspeech.2011-771

Levitan, R., Beňuš, Š., Gálvez, R.H., Gravano, A., Savoretti, F., Trnka, M., Weise, A., & Hirschberg, J. (2016). Implementing acoustic-prosodic entrainment in a conversational avatar. *Interspeech 2016,* 1166–1170. San Francisco, USA. https://doi.org/10.21437/Interspeech.2016-985

Lombard, É. (1911). Le signe de l'élévation de la voix, *Annales des Maladies de L'Oreille et du Larynx*, *37*(2), 101–119.

Lu, Y. & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America, 124*(5), 3261–3275. https://doi.org/10.1121/1.2990705

Lu, Y. & Cooke, M. (2009a). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *The Journal of the Acoustical Society of America, 126*(3), 1495–1499. https://doi.org/10.1121/1.3179668

Lu, Y. & Cooke, M. (2009b). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication, 51*, 1253–1262. https://doi.org/10.1016/j.specom.2009.07.002

Marxer, R., Barker, J., Alghamdi, N., & Maddock, S. (2018). The impact of the Lombard effect on audio and visual speech recognition systems. *Speech Communication, 100*, 58–68. https://doi.org/10.1016/j.specom.2018.04.006

Nonaka, S., Takahashi, R., Enomoto, K., Katada, A., Unno, T. (1997). Lombard reflex during PAG-induced vocalization in decerebrate cats. *Neuroscience Research, 29*(4), 283–9. https://doi.org/10.1016/s0168-0102(97)00097-7

Natale, M. (1795). Convergence of Mean Vocal Intensity in Dyadic Communication as a Function of Social Desirability. *Journal of Personality and Social Psychology, 32*(5), 790–804. https://doi.org/10.1037/0022-3514.32.5.790

Pardo, J.S., Gibbons, R., Suppes, A., Krauss, R.M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*, 190–197. https://doi.org/10.1016/j.wocn.2011.10.001

Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R., & Kearney, J. K. (1989). Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America, 85*, 894–900. https://doi.org/10.1121/1.397561

Raitio, T., Suni, A., Pohjalainen, J., Airaksinen, M., Vainio, M., Alku, P. (2013). Analysis and Synthesis of Shouted Speech. *Interspeech 2013*, 1544–1548. Lyon, France. https://doi.org/10.21437/Interspeech.2013-391

Rostolland, D. (1982a). Acoustic features of shouted voice. *Acustica, 50*(2), 118–125.

Rostolland, D. (1982b). Phonetic structure of shouted voice. *Acustica, 51*(2), 80–89.

RStudio Team (2020). RStudio: Integrated Development for R. (computer program) RStudio, PBC, Boston, MA. https://www.rstudio.com/

Šimko, J., Beňuš, Š. & Vainio, M. (2014). Hyperarticulation in Lombard speech: A preliminary study. *Proceedings of the 7th International Conference on Speech Prosody*, 869–873. https://doi.org/10.21437/SpeechProsody.2014-162

Šimko, J., Beňuš, Š. & Vainio, M. (2016). Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue. *The Journal of the Acoustical Society of America, 139*, 151–162. https://doi.org/10.1121/1.4939495

Šimko, J., Vainio, M., & Suni, A. (2020). Analysis of speech prosody using WaveNet embeddings: The Lombard effect. *Proceedings of the 10th International Conference on Speech Prosody*, 910–914. https://doi.org/10.21437/SpeechProsody.2020-186

Uemura, Y., Morise, M., & Nishiura, T. (2010). The Lombard speech recognition based on the voice conversion towards neutral speech. *Proceedings of 20th International Congress on Acoustics, ICA 2010, PaperID, 167*.

Van Heusden, E., Plomp, R. & Pols, L.C.W. (1979). Effect of ambient noise on the vocal output and the preferred listening level of conversational speech. *Applied Acoustics, 12*(1), 31–43. https://doi.org/10.1016/0003-682X(79)90037-9

Wagner, P., Trouvain, J., & Zimmerer, F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics, 48*, 1–12. https://doi.org/10.1016/j.wocn.2014.11.001

Xue, Y., Marxen, M., Akagi, M., & Birkholz, P. (2021). Acoustic and articulatory analysis and synthesis of shouted vowels. *Computer Speech & Language, 66*(2), 101156. https://doi.org/10.1016/j.csl.2020.101156

Yanushevskaya, I., Gobl, C., & Ní Chasaide, A. (2013). Voice quality in affect cueing: does loudness matter? *Frontiers in Psychology, 4*, 335–335. https://doi.org/10.3389/fpsyg.2013.00335