The Lombard effect

THE EFFECTS OF NOISE EXPOSURE AND BEING INSTRUCTED TO SPEAK CLEARLY ON SPEECH ACOUSTIC PARAMETERS

Jessie Waaning

Radboud University Nijmegen

Master's thesis of General Linguistics

Supervisor: Dr. Esther Janse

19 July 2021

Index

1. Introduction

Whenever the hearing ear of a unilaterally deaf person is covered with an active noise-generating device, their speech changes from 'natural' speech as we all recognize it, to the type exhibited by people who are deaf. Étienne Lombard discovered this effect while working with unilaterally deaf patients in 1911. After testing some observations through a set of experiments, Lombard concluded that hearing people raise the volume of their voice abnormally when speaking in a noisy environment. He postulated that the volume increase was necessary for both making oneself (more) intelligible, as well as to hear oneself better (Lombard 1911). This effect would become known as the Lombard effect, also referred to as 'Lombard sign', 'Lombard reflex' or 'Lombard speech', all after its namesake (Zollinger and Brumm, 2011). The Lombard effect is, in essence, speech affected by background noise. Extensive study of the phenomenon has revealed a number of particular differences between Lombard and natural speech. Studies consistently find that Lombard speech has, for example. higher vocal intensity and fundamental frequency than normal speech. Despite the breadth of literature, studies are limited on the question whether, and if so, how the Lombard effect changes as speakers continue to speak in noisy environments. Change here indicates whether the differences found between Lombard and normal speech become larger or decrease. Take for example the question whether Lombard speech's increased vocal intensity grows over time. Learning whether the Lombard effect changes as people speak, will improve our understanding of it, and pave the path for future research on the phenomenon. The aim of this thesis will be to contribute to Lombard effect research, by studying whether Lombard speech strategies change as participants use Lombard speech, in comparison to them using natural speech. Lombard speech studied in this thesis includes instructions for so-called 'clear speech'. I will start by defining this and other terms below, while also identifying the scope and circumstances of my and past research. This work's formal research question will be discussed at the end of the introduction paragraph.

1.1. Defining the Lombard effect

The Lombard effect is an alteration in speech parameters, found in speakers who are speaking in a noisy environment. Its properties can vary as a result of several influences, including noise type, speaker difference, and even type of sentences produced. Despite the large variety of possible influences, the noted shift in parameters that occurs when Lombard speech is

used is largely consistent over the literature. The next paragraph will deal with these in detail. Furthermore, the nature of Lombard speech will be discussed: there is the question of whether the Lombard effect is reflexive, whether the Lombard effect occurs as an unconscious reaction (a reflex) to a noisy environment, or whether the phenomenon is at least partially conscious and done through volition. The literature is as of yet inconclusive on that question, with the literature showing mixed results. This debate will discussed in paragraph 1.1.3.

1.1.1. Lombard speech parameters

Lombard speech, compared to natural speech, is characterized by a change in the following parameters:

- **increased vocal intensity**, which is the loudness of the speech produced;
- **increased fundamental frequency (F0)**, which is the pitch of the speech produced (e.g. Summers, Pisoni, Bernacki, Perlow, & Stokes, 1988; Junqua, 1993 and 1996; Lau, 2008; Lu & Cook, 2009; Folk & Schiel, 2011; Garnier & Henrich, 2014);
- **shifted center frequencies of vowels for F1 and F2**, F1 and F2 frequencies directly representing the position of the tongue during pronunciation, height and anteriority respectively (Summers et al., 1988; Junqua & Anglade, 1990; Junqua 1993; Garnier & Henrich, 2013);
- **increased vowel and/or word duration**, which is directly linked to rate of speech (Junqua, 1993; Lau, 2008; Garnier, Bailly, Dohen, & Loevenbruck, 2006; Garnier & Henrich, 2014);
- **flatter spectral tilt**, spectral tilt being the distribution of energy at different frequencies (Summers et al., 1988; Lu & Cooke, 2009).

These characteristics correspond to noticeable changes in speech. The following manifestations are the result of the altered parameters: louder speech, higher pitched speech, different articulation, and slightly more shouted speech. Increased intensity makes that Lombard speech is louder than natural speech. The increase in F0 is audible as higher pitched speech under Lombard conditions. The shift of formant center frequencies for F1 and F2 in vowels means that the average F1 and F2 of these vowels manifest at different frequencies in Lombard speech than in natural speech. This shift indicates a difference in articulation style. If the formant center frequencies of vowels move further away from other vowels and thus increase the vowel triangle

of a speaker, that indicates hyper-articulation: individual vowels become more pronounced. The increase in vowel and/or word duration indicates a lower rate of speech. In Lombard speech, more time is taken to pronounce words. The last change in speech is realized by the flattening of spectral tilt, which means that the distribution of energy, measured in dB, becomes more equal over different frequencies. (Summers et al., 1988; Lu & Cooke, 2009). In natural speech, low frequencies receive more spectral energy than high frequencies, while in Lombard speech the energy spread is evened out. Garnier and Henrich (2014) find that energy in Lombard speech is boosted specifically at the frequency where the ear is most sensitive (around 3-4 kHz). This coincides with spectral tilt flattening, as a flatter spectral tilt indicates more energy at higher frequency levels of speech (Lu & Cooke, 2008). The reader may note that Lombard speech parameters bear a resemblance to those associated with shouted speech. Shouted speech is also characterized by an increased intensity, and a flattened spectral tilt is an exemplary characteristic of shouted (Raitio, Suni, Pohjalainen, Airaksinen, Vaino & Alku, 2013). However, it is to be noted that Lombard speech differs from shouted speech in two fundamental ways: the extremity of intensity, and intelligibility. First, shouted speech "lies at the extreme end of the vocal effort continuum" (Raitio et al., 2013, p.1). This means that it is produced at the maximum level of a speaker's vocal intensity. Lombard speech however, is produced way below this maximum intensity level. Second, shouted speech is generally measured as being less intelligible than natural speech, while studies show that Lombard speech is more intelligible than natural speech. So, although Lombard speech and shouted speech share similar parameters, their difference in intensity and intelligibility make them different phenomena. That last point raises the following question: why is Lombard speech considered more intelligible than natural speech, while shouted speech is not? It is unlikely that this is merely the result of difference in intensity. Indeed, it seems that the exact combination of Lombard speech parameters is responsible for this. After manipulating various individual Lombard speech characteristics, Lu and Cooke (2009) find that the particular combination of factors contribute to the increased intelligibility of Lombard speech in noisy environments. Spectral tilt flattening explains some, but not all of Lombard speech's increased intelligibility. While an increase of F0 by itself did not significantly change intelligibility, it did improve total intelligibility when combined with spectral tilt flattening. In conclusion, aforementioned factors ensure that Lombard speech has superior intelligibility over natural speech in challenging environments (Summers et al., 1988; Garnier & Henrich, 2014).

The sum of these parameter changes in the context of speaking in a noisy environment, is what defines the Lombard effect.

1.1.2. Variability within the Lombard effect

The *direction* of changes in acoustic parameters, such as that intensity increases rather than decreases, is largely consistent over the literature. But there is variation in the *degree* to which these changes occur, as many variables can influence the manifestation of the Lombard effect. These variables include, but are not limited to:

- **Types of masking noise**: the noise used so speakers do not hear themselves, and can exist on a spectrum ranging from speech noise to white noise;
- **Speaker specifics:** the manner in which speakers differ from one another. The most important property for this thesis is speaker sex;
- **Various components of the produced sentences:** include such as language variation, vocabulary size, and measured vowels.

These variables considerably influence the manifestation of Lombard speech. They therefore need to be addressed and taken into account when studying the effect. Below, I will go into this thesis's methods for doing so.

Firstly, in order to study the Lombard effect, a so-called masking noise must be generated so that speakers' ability to hear themselves is nullified or vastly diminished: their auditory input is filled with data that has no semantic value. In most experiments, masking noise consists either of white-Gaussian noise (e.g. Summers et al., 1988), cocktail party noise (also: multi-talker environment) (e.g. Junqua 1996) or speech-shaped noise (e.g. Lu and Cooke, 2009). White-Gaussian noise produces a sharp sound with a spectrum similar to that of an /s/, ensuring a uniform noise intensity across all frequencies. Cocktail party noise uses a multitude of speakers of varying intensity as masking noise, resulting in noise that competes with a speaker's own speech frequencies. Speech-shaped noise consists of the same spectral distributions as speech, without using actual speech. Its resulting frequencies center around 2-4 kHz. Speech-shaped noise is similar to multi-talker environments, but with artificial noise rather than actual speech being generated, allowing for a constant intensity. In contrast to multi-talker noise, speech-shaped noise does not contain any phonetic or lexical information in its masking (Varnel, Meyer, Hoen, & Meunier, 2012). The different masking noises have varying effects on Lombard speech, with certain articulatory and acoustic parameters such as lip movement and F0, increasing more

in certain types of noise than in others (Garnier et al., 2006; Garnier & Henrich, 2014). The direction of the these parameters' alteration patterns is similar over all masking noises. The Lombard effect seems to occur specifically if there is frequency competition between masking noise and speech, with masking frequency being in the same range as speech – between the 0.05 and 4.0 kHz (Stowe and Golob, 2013). How Lombard speech manifests, is thus firstly dependent on the type of masking noise used. For the purposes of this thesis's study and the interpretation of its results, it should be noted that the Lombard speech data used in this thesis was recorded with speech-shaped noise. In the corpus used for this study, the Radboud Lombard Corpus (RaLoCo), Chen and Janse (2019) chose to use speech-shaped masking noise because it simulates speech, while allowing a level of control over intensity, and because of its masking frequency. Its frequency centers around that of speech, giving it a considerable chance to evoke the Lombard effect.

Secondly, the Lombard effect manifests itself differently depending on the individual speaker performing it. Differences between people ensure that they produce speech at varying levels of intensity, at different pitch, and so on. Physiological properties, such as length of the larynx, affect speech production. Of the many possible physiological properties that affect Lombard speech, speaker sex seems to be the one most studied in the literature. Perhaps this is because speaker sex often entails a radical shift in physiological properties associated with the voice, as a result of the effects of puberty – men tend to develop deeper voices. Whatever the case, speaker sex is the factor that has the most data available about it in the literature, so that this paragraph will focus on it and its implications. With its effect on physiological properties, speaker sex is a reliable variable for Lombard speech variance. Speaker sex influences the extent to which a speaker's speech parameters change, and with that, the intelligibility of a speaker's Lombard speech. In a study set up to improve automatic speech recognition, Junqua (1993) finds that when speaking in a multi-talker environment, female speakers tend to be more intelligible than male speakers. The opposite is found to be true for normal speech. A study by Garnier and Henrich (2014) provides a possible explanation for the reduced intelligibility of male speakers in a noisy environment: men tend to adjust their speech to a lesser extent. In their study, they measured multiple characteristics of Lombard speech and looked at the differences between male and female speakers, among other things. They find that men and women alter their speech in similar manners, such as a shifting frequency or F1 in vowels. However, this shift is smaller in

men's speech than women's. The resulting speech parameters then differ between men and women, with men's speech parameters conflicting with the acoustic parameters of multi-talker environments. This is essentially what makes men less intelligible, providing an explanation as to why Junqua (1993) finds Lombard speech produced by men to be less intelligible than Lombard speech produced by women. Additionally, speaker sex differences could be culturally constructed (Pépiot, 2014), but further investigation is necessary to determine to what extent cultural influences affect Lombard speech parameters. In order to find out whether Lombard speech techniques change over time, speaker sex differences must be taken into account. To this end, a random speaker effect is implemented in the calculations. For the purposes of this study, speaker sex is not taken into account in the final calculations, as this was out of scope for this thesis.

Thirdly, various components of the produced sentences themselves can influence the manifestation of Lombard speech. Dependent on language type (Junqua, 1996), vocabulary size (Junqua 1993), and measured vowels (Garnier & Henrich, 2014), Lombard speech can manifest differently. Language type influences the manifestation of Lombard speech, with vowel shifts occurring in some languages, and not in others. Comparing Lombard speech studies across languages, Junqua (1996) finds that acoustic differences between Lombard and natural speech across American English, French, Spanish, and Japanese was similar in general, with varying degrees of change between the languages. For example, the F2 increase of vowels is more prominent and consistent in Spanish Lombard speech than in American English and French Lombard speech. Another comparison by Junqua shows that Japanese Lombard speech has F1 and F2 shifts that are not found in American English. In Japanese Lombard speech, F1 and F2 values increase for vowels with base frequencies below 1500 Hz, and shift to lower frequencies for the vowels with formant frequencies above 1500 Hz. This is not found in the American English Lombard speech studies used in the comparison. However, the study by Summers et al. (1988) does show a formant shift tendency that represents the formant shift in the reported Japanese study: although not significant, Summers et al. find a trend of vowels containing high F2 frequencies decreasing in formant frequency in Lombard speech, and vowels of low F2 frequencies increasing in the Lombard condition. The Summers et al. study uses English speakers, presumably American, and was not used in the comparison by Junqua (1996). Nevertheless, the tendency of vowel formant shift was significantly found in the Japanese

studies, while not reaching significance in the (American) English studies, showing a difference between languages all the same. This thesis uses native Dutch speaker, and its results can be used to improve our understanding of the effects of language type on Lombard speech manifestation. Vocabulary size influences the manifestation of Lombard speech, as the more words are used in a sentence, the less intelligible Lombard speech becomes. Junqua (1993) tests the influence of vocabulary size on normal and Lombard speech production by having listeners judge the intelligibility of sentences produced under noisy conditions. Junqua does not define 'intelligibility' (for instance in terms of acoustic parameters), but relies on the innate lexical judgment of native English listeners. The study shows that intelligibility of Lombard speech decreases as vocabulary increases, meaning that the more words are used in a Lombard speech sentence, the less intelligible the utterance becomes. This information will be taken into account when performing tests with Lombard speech, as to better interpret the results of this thesis: after all, not all stimuli used between and in various studies will have the same amount of words. Measured vowels influence the manifestation of Lombard speech, as individual vowels vary in which formants show an increased frequency. Garnier and Henrich (2014) measure speech parameters of vowels as one part of their study on intelligibility in Lombard speech. Their results show that different vowels change differently when used in a Lombard condition, as compared to a natural condition. Their measured vowels, /a:/, /i:/, and /u:/, all show increases in formant frequencies. However, for each vowel, different formants are subject to this increase. For female speakers, vowel /a:/ shows a clustered increase of the formants F2 and F3. Vowel /i:/ shows two different tendencies across participants, with a clustered increase of the F2 and F3 formants and a cluster of the F4 and F5 formants in one group of people, and one F2 F3 F4 cluster in the other. Vowel /u:/ shows an F3 peak, and a clustered F4 and F5 peak. The vowels also change differently depending on speaker sex and background noise, resulting in complex interactions for which I kindly refer to their study. The variation between vowels is likely due to the initial frequencies of these vowels: some studies mentioned in the literature analysis by Junqua (1996) show that vowels with a base frequency lower than 1500 Hz change differently in a noisy environment than those with a frequency above 1500 Hz. Garnier and Henrich (2014) attribute this difference to an increase in intelligibility: although the main strategy of speakers in their study is to adapt vocal effort to the intensity of background noise, the specific differences in change between vowels indicate an adaptation to improve speech intelligibility. The current

thesis includes a vowel analysis that focusses on the same vowels, /a:/, /i:/, and /u:/, because they are the corner vowels of the Dutch language. Differences between vowels will thus be best captured by these three vowels.

In summary, many components influence the manifestation of the Lombard effect. This work seeks to take inventory of these components, and take careful decisions concerning which components to use or exclude. In order to create reliable and sufficiently actionable data, this thesis uses speech data produced with speech shaped noise, and takes speaker differences and sentence variability into account. Speaker sex is not taken into account in the final calculations, as it is considered out of scope for this thesis, as it would have made calculations excessively complex. As speaker sex is a known influence on variability of the Lombard effect, the results as presented in this thesis may slightly deviate from the actual realization of speech collected in the RaLoCo.

1.1.3. Nature of the Lombard effect

Lombard believed his patient to be unaware of the change in the latter's speech, which is why the effect has often been referred to as a reflex (Zollinger and Brumm, 2011). Whether the Lombard effect is a reflex or a conscious and willed act is still a matter of debate. Additionally, it is difficult to fully differentiate between Lombard speech and another speech phenomenon: clear speech. Below, readers will find elaborations on the debate of 'Lombard-as-reflex' and 'Lombard-as-act', as well as an elaborate discussion of the grey area between Lombard and clear speech. It must be stated however, that this work does not seek to settle this particular contention once and for all. This work will focus solely on elucidating the results of its own particular subject and scope. This work will still make sure to note whenever the results can contribute to the contemporary discussions. In doing so, this thesis seeks to contribute (albeit passively) to these branches of Lombard speech study, in hopes that it might help lead to advancements in related fields.

A reflex is an automatic reaction to a stimulus, and reactions to sensory stimuli are standard responses (Pomfrett, 2005). Lombard-as-reflex should occur without the active intent or conscious awareness of the speaker. Most research on Lombard speech is conducted with the speaker reading out sentences in a natural or a noisy environment (e.g. Stowe & Golob, 2007). In these experiments, isolated speakers produce Lombard speech. There are no interlocutors, nor are speakers told to pay attention to speech. For many researchers, the fact that Lombard speech

is produced under these conditions confirms that Lombard speech is 'reflexive' in nature, as the Lombard effect seems to take place automatically, without the speakers' conscious awareness or volition playing a role.

Dissimilarly, other researchers consider Lombard-as-act to be a better descriptor of noted results. While not denying that some manner of reflexive behavior is likely occurring in Lombard speakers, researchers like Tuomainen and Hazan (2018) believe that volition on the part of speakers also plays a role. Their reasoning is as follows: speakers will try their best to make themselves intelligible to their listeners, even when not prompted to do so. As a result, they will willfully and consciously attempt to speak more intelligible than they would under regular circumstances. Some Lombard speech studies even include instructions to "speak clearly" when in the noisy test-environment (e.g. Summers et al., 1988). In Tuomainen and Hazan's interpretation, Lombard speech overlaps with so-called 'clear speech', a term used to describe the acoustic-phonological changes speakers make to increase their intelligibility in challenging communication conditions. The speakers' attempt to be more intelligible would thus suggest that Lombard speech does not occur (only) as a reflex, but is more likely to be a conscious act.

One cannot help but notice the apparent overlap in definitions between Lombard and so-called 'clear speech'. The literature is not yet fully clear on the exact distinction between these two. Lee and Baese-Berk (2020) define clear speech as "a speaking style that speakers adopt when they talk with listeners who they anticipate may have a problem understanding their speech" (3702). This includes a large variety of situations, such as foreign directed speech, as in Lee and Baese-Berk (2020), or hearing impaired speech, as in Picheny, Durlach, and Braida (1986). In this latter view, clear speech can also occur when two people communicate in a noisy environment, and the speaker anticipates that environmental noise will interfere with the message delivery. Speech produced in this situation could also be identified as Lombard speech, as its definition consists of the same modifications to speech occurring. Thus, the definitions of Lombard and clear speech overlap. Not only do the definitions of Lombard and clear speech overlap, they share similar phonetic characteristics. Clear speech, in contrast to natural speech, is characterized by: an increased F0, a slower speaking rate, and a shift of spectral energy to the higher frequencies (Krause & Brauda, 2004; Hazan & Baker, 2011). These are generally observed phenomena, as individuals may adopt divergent strategies to achieve the same goal (Krause & Brauda, 2004). All of these characteristics are also found in Lombard speech. Further

complicating matters is the evidence that some Lombard studies include instructions for participants to "speak clearly". The result is two speech phenomena whose ostensibly separate study, has become rather complexly entangled.

In an attempt to partially untangle the two, Cooke, King, Garnier, and Aubanel (2014) distinguish interlocutor and environment related speech modifications. In their interpretation, interlocutor-directed speech consists of speech modified to a listener, whom the speaker perceives to have intrinsically reduced audibility, regardless of environment. This definition most readily demarcates clear speech for the purposes of study. Environment-derived speech modifications on the other hand, occur when audibility, including self-audibility, is affected negatively by noise. That is: when some environmental property lowers the capacity of a listener to hear their own or another's message. This definition encompasses Lombard speech. As a general rule, clear speech can thus be seen as a listener-oriented phenomenon, while Lombard speech is an environment-derived phenomenon. This mechanism of demarcating between the two phenomena is reflected in the manner in which studies on the two subjects are conducted: clear speech studies often exclude noisy environments for the speaker while they are speaking. For example, to elicit clear speech in a conversational setting, Hazan and Baker (2010, 2012) added noise to the speech being received by a participating speaker, suggesting that the listener will be in a noisy environment, while the speaker themselves was situated in a silent (i.e. non-noisy) environment as they produced speech. Similarly, Lombard speech research is most often carried out without a direct conversational partner, in order to exclude listener-oriented effects. The data used for this thesis is an example of this latter setup: it uses a setting where speakers are exposed to noise over headphones, while alone in a booth (Shen & Janse, 2019). Presumably, this mechanism of demarcation between Lombard and clear speech is applied to improve the quality of the data produced and avoid interpretational difficulties stemming from the overlap between two potentially very different phenomena. Nevertheless, Cooke et al. (2014) emphasize that, although they attempt to distinguish the two, their efforts are unsatisfying: ultimately, the definitional issue requires more study to be fully understood. As far as I know, no studies exist that provide conclusive evidence for the absence of environment-derived aspects in clear speech, nor are there studies that provide evidence of the absence of listener-oriented aspects in Lombard speech. As an additional complicating factor, a recent study suggests that clear speech may not be fully listener-oriented, but can become speaker-driven in time. Lee and Baese-Berk (2020) set

out to measure the listener-oriented nature of clear speech. In their study, they use Baker and Hazan's (2010) LUCID corpus on clear speech in foreigner directed speech, which is known to elicit the clear speech phenomenon. In the corpus, native English speakers carry out what is called a Diapix task with non-native English speakers. The Diapix task is a problem-solving puzzle used to elicit spontaneous speech between a pair of talkers. Two speakers are presented with nearly identical pictures, each speaker only being able to see their own picture. The speakers are asked to figure out the differences between their two pictures by speaking to each other. In the corpus, Native English speakers are paired with non-native English speakers of various backgrounds. They perform the Diapix test three times, meaning they discuss three different pictures. Participants do not receive instructions to "speak clearly". Lee and Baese Berk (2020) confirm that the speech from the LUCID corpus is clear speech, through acoustic analysis. For the main experiment, speech from the LUCID corpus is first rated on intelligibility, as intelligibility is one of the main characteristics of clear speech. This intelligibility rating is the main topic of their analysis. Their analysis shows that native English speakers use speech that is generally more intelligible than normal speech over all Diapix tasks, confirming their use of clear speech. However, in the time period of discussing one picture, Lee and Baese-Berk find that intelligibility of the native English speakers is inconsistent within the tasks. Speakers use highly intelligible speech at the start of each Diapix task. However, as the conversation about one picture progresses, intelligibility decreases and continues to do so. When the conversation about one picture ends, and the next is presented, speech intelligibility increases again. Lee and Baese-Berk interpret this as a 'reset' of participants' speech. The change in intelligibility, they theorize, could be a result of a change in what drives the speech. The *presence* of clear speech may be listener driven in this case, which is signified by the high intelligible type of speech, which is easier for the non-native English speaker to understand. The further *maintenance* of clear speech however, may be driven by the speaker, for whom highly intelligible speech can be effortful. The decrease in intelligibility within one picture discussion could thus be a result of the speaker prioritizing their own effort over the listener's effort. Once a new picture is introduced to be discussed, clear speech is re-instigated and the speaker prioritizes the listener again, hence the 'reset'. In summary, clear speech may not be only listener-driven, but also be or become speaker-driven. Lee and Baese-Berk's findings therefore further complicate a clear demarcation of Lombard speech as distinct from clear speech. Future studies on the listener-, speaker-, and

environment-derived factors of both types of speech could shed more light on a potential solution to this problem. However, at the time of thesis's writing, Lombard and clear speech continue to have significant definitional overlap. Therefore, I will treat Lombard speech as described in Cooke et al. (2014): as an environmental phenomenon, that occurs when spoken in noise and has overlap with clear speech.

The corpus used in this thesis, like previous studies on Lombard speech, includes instructions for the participants to speak as clearly as possible. Consequently, it overlaps with clear speech studies. However, providing a solution for the ambiguity of the nature of Lombard speech is not the topic of this thesis. By concerning itself with speech that is produced in noisy environments, and excluding a direct interlocutor when participants speak, this thesis will suffice in seeking to preserve a continuation of previous Lombard speech research. Regardless of the definitional overlap as explored, this thesis's experimental results are suitable for expanding Lombard speech literature, if for the mere fact that they attempt to replicate and expand on previous findings.

1.1.4. Conclusions about the Lombard effect

The Lombard effect can be summarized as: an alteration in speech that occurs when a speaker is in a noisy environment. The phenomenon is flexible, as many variables influence its manifestation. Lombard speech is speaker-specific, dependent on the type and frequency of interrupting noise, and on the content of the target message. The strategies used to create the Lombard effect nevertheless include the same directional shift in acoustic characters, albeit to different extents. The literature is inconclusive about whether the Lombard effect is reflexive or not. The same question haunts the exact demarcation of Lombard speech from clear speech. Understanding how Lombard speech techniques change over time will add to the understanding of the nature of Lombard speech, and encourages research on speech-related subjects to include usage-effects.

1.2. Research Question

At the time of writing, there are no studies on how Lombard speech parameters change as a person uses Lombard speech. In order to improve our understanding of Lombard speech's nature, I will be researching the effects of using Lombard speech on speech parameter changes, and how this differs from speech produced under natural speaking conditions. By calculating the changes participants show from trial to trial, I hope to reveal more of Lombard speech's

workings. This will not only further understanding of the effect itself, but hopefully provide insight into the adaptability and plasticity of speech behavior. Therefore, the research question of this thesis can be posed as thus:

> RQ: How do acoustic speech parameters change over time spoken, and does this change differ for Lombard speech when compared to natural speech?

This thesis tries to capture whether acoustic speech parameters change as speakers continue to speak, and whether the potentially resulting speech parameter alteration differs between Lombard and natural speech. To this end, speech in natural environments is compared to speech in a noisy environment, at different points of time spoken. In short: how do the properties of speech change over time (of continuous use), in quiet and noisy environments? Previous research on Lombard speech has defined the concept through determining the parameters that make up the effect. With this thesis I hope to further increase Lombard speech's understanding, by studying previously unstudied effects of using of Lombard speech, on speech parameters. Finally, I hope to inspire further research on (short-term) usage effects on other speech phenomena, as this is a relatively unexplored theme.

To the end of measuring the effects of prolonged use of Lombard speech, I will study the change in speech parameters as a function of time in both natural and Lombard speech, and measure the difference between Lombard and natural speech parameters. The change of speech parameters as a function of time will be dubbed 'the trial effect' for the purposes of this thesis. Measuring these trial effects has three possible outcomes: the distance in acoustic-phonological adaptions between natural and Lombard speech could decrease, increase, or remain stable over trials. These three hypotheses will be elaborated upon below.

Firstly, a decrease in Lombard speech acoustic-phonological adaptions could result from the experiments. One possible explanation for the 'decrease-hypothesis' could be speaker fatigue. The increased intensity of Lombard speech and noise exposure could affect the speaker's vocal effort, leaving them fatigued. In researching the effects of reading aloud for long durations, Gelfer, Andrews, and Schmidt (1991) find that participants who had received no professional vocal training anterior to the experiment, experienced negative speaking voice effects after an hour of reading at higher intensity levels. In a study on the influence of speech intensity and

speed of reading aloud on oxygen consumption, Moon and Lindblom (2003) find that with both higher intensity and higher rate of speech, oxygen consumption also increased. The researchers link increased oxygen consumption to higher physical strain of the speaker. Lombard speech consists of a higher intensity, but a lower rate of speech than natural speech. Therefore we cannot predict how oxygen consumption on its own will affect the manifestation of the Lombard effect. However, both the Gelfer et al. (1991) and Moon and Lindblom (2003) studies link high intensity speech to increased vocal effort, likely resulting in vocal strain. As higher intensity is one of Lombard speech's acoustic parameters, it is possible that Lombard speech also takes a physical toll on the speaker's vocal cords. In addition to the influence of higher spoken intensity, Kristiansen, Lund, Persson, Shibuya, Nielsen, and Scholz (2014) find that vocal fatigue symptoms also correlate positively with average noise exposure. In their study, they measure the average noise exposure and speaking time of teachers, and its effects on vocal and mental fatigue. They find that this correlation between noise exposure and fatigue is even more significant than the correlation between vocal symptoms and average vocal load. Kristiansen et al. also find that noise exposure increases the experience of 'fatigue in the head', a measurement on the self-report form that teachers filled in for the study. The study shows no correlation between noise exposure and stress scores, however. Their study shows that average noise exposure significantly increases vocal and mental fatigue, but not stress. The idea that noisy environments affect cognitive functioning is also shown by Rotton, Olszewski, Charleton, and Soler (1978). In their study, participants are exposed to (speech and non-speech) noise while performing recall and differentiation tasks. The results of Rotton et al.'s study show that background noise reduces participants' ability to tolerate frustration and to differentiate among people. Combining the effects of loud speech on vocal strain with the physical and mental fatigue of being in noisy environment results in a high possibility of fatigue caused by Lombard speech, itself leading to a decrease in speech adaptions. An example of decrease in a speech parameter can be seen in the results of Lee and Baese-Berk's (2020) study, where intelligibility (which signified clear speech) decreases within the discussion of one Diapix task. Although intelligibility did not vary in all of the early stages of image discussion, intelligibility decreased in later stages of the process. As explored previously, Lombard speech is closely related to clear speech, and both share intelligibility as a parameter. Lee and Baese-Berk also believe that the decrease in intelligibility is the result of clear speech becoming less interlocutor-driven, and

more speaker-driven. We can therefore carefully compare Lee and Baese-Berk's study to the current thesis. If a speaker becomes fatigued, and speech is speaker-driven (and/or not interlocutor-driven), a decrease in fatigue-inducing speech parameters is to be expected. As Lombard speech is taken to be environmentally-driven rather than interlocutor-driven, it is possible that the results of the current study will replicate this aspect of Lee and Baese-Berk's results, in which speakers decrease their effort in the latter half of image discussion. In that case, a decrease of acoustic-phonological adaptions would be measured.

Secondly, an increase in Lombard speech acoustic-phonological adaptions could result from the experiments. One possible explanation for this 'increase-hypothesis' could be that a training effect is taking place. Speakers could become better at realizing the Lombard parameters as they practice using Lombard speech. Increasing F0 and intensity could for example become easier to effect over time for speakers. There are several factors that this adaption could measurably manifest in: growing accustomed to the use of one's voice at higher intensity, growing accustomed to the environment, adjusting one's speaking speed, etc. If speakers were to become accustomed to speaking in a noisy environment, that would allow them to further focus on producing Lombard speech. In Gelfer, et al. (1991) study, participants trained as singers with vocal control training showed an increase in vocal quality after reading aloud for one hour, measured by a decrease in negative vocal qualities such as jitter (see their study for a more detailed description). This suggests that their vocal performance improved over the duration of their task. However, this thesis's participants have not received any vocal training, nor were they selected for any such property or experience. As previously mentioned, participants in the Gelfer et al. study who had no previous vocal training, contrastingly did not improve vocally after the task, but rather, showed measurable and increasing strain on their voices. It must be noted at this point however, that the Gelfer et al. testing phase took considerably longer than the one used for this thesis. Whereas this thesis's duration is less than 10 minutes, Gelfer et al.'s testing took about an hour each time, for each participant. This thesis's test is thus far less intensive, as it requires less than ten minutes of speaking time in total. Additionally, Gelfer et al. measured changes in participants' performance at only two instances: one time before the experiment, and one time after an hour of reading. This means that we have no fine-grained data about how Gelfer et al.'s participants' vocal strain developed over that period. As a result, there is no way to tell if for instance, a training effect preceded the strain and decrease in quality. Another factor

that could contribute to Lombard speech becoming easier to produce, is the influence of oxygen consumption on vocal strain. Lombard speech is further defined by a lower speaking rate. Moon and Lindblom (2003) tie that property to decreased oxygen consumption, which should in turn translate to decreased vocal strain. However, combined with the increased oxygen consumption that occurs with an increased intensity, it is difficult to predict whether total oxygen consumption decreases or increases, and thus, how oxygen consumption will affect speakers' vocal strain. It is possible that having to adapt the same acoustic parameters of one's speech continuously could make it increasingly easier to replicate the desired effect. As a person gets used to the challenging environment, they might adapt their speech to a form that is most effective in the given situation. If the results of this thesis show an increased distance between natural and Lombard speech parameters this could be an indication that participants have improved their Lombard speech while they were using it. That would in turn provide evidence for the 'increase-hypothesis'.

In the third and final hypothesis, speech parameters could prove stable over time, where no overall change in adaptions of speech parameters would be discernible. Finding these result would be more complex to interpret: stable parameters could be the result if Lombard speech is a reflex, as reflexes do not change over time. Additionally, finding stable speech parameters could be the result of speech produced in a stable environment. However, not finding any change could also be the result of a flawed methodology. Stable parameters can be the result if there is no change in the underlying motivational drivers for speech, especially if speech produced is a result of a reflex. The issue of Lombard-as-reflex or Lombard-as-act is thus important for the interpretation of our results here. If our results show no change in Lombard speech parameters over time, this could be interpreted as evidence for Lombard-as-reflex. Lombard speech is thought to be environment driven, so if the environment does not change, it is possible that produced speech will not change over time. Lee and Baese-Berk (2020) did groundbreaking work in studying the effects of continuous use and presence of a speech phenomenon, on the manifestation of said phenomenon. Their participants performed the Diapix task three times. In addition to the previously discussed results of their study, Lee and Baese-Berk find that the level of intelligibility in the early stages of Diapix tasks remains constant over all tasks. These early stages, which have the highest intelligibility, are considered the 'interlocutor-driven' aspects of the conversation. Lee and Baese-Berk consider these early stages the 'reset' points of the

participants' speech. Speakers' willingness to engage in or maintain clear speech decreases over time. As previously mentioned however, Lombard speech is defined as environment-driven. As long as the drive for the type of speech does not change, it is possible that the speech itself will also remain constant. That would prove to be the case even more so if Lombard speech is a reflex: a largely involuntary reaction to a particular environment. If the Lombard effect is indeed a reflex elicited by noisy environments, speakers will not consciously and consistently adapt their technique: reflexes manifest uniformly within individuals when regarding parameters such as type of noise, etc. (Pomfrett, 2005). If the results show a difference in parameters between Lombard and natural speech, but no change of these parameters over time (so the parameters remain constant) this could thus be an effect of Lombard-as-reflex. Alternative outcomes or explanations to the ones described for the third hypothesis, remain a possibility. If no change in acoustic speech parameters is recorded, that does not necessarily imply a training effect or provide evidence for Lombard-as-reflex. Instead, it could just as well be that both normal and Lombard speech show the same pattern of change over trials, resulting in their differences remaining stable – in turn complicating data interpretation for the purposes of the research question. Alternatively, this study's sample size might turn out to be too meager to produce interpretable results on this point. A larger dataset might tease out results that only become visible at a particular critical mass of participants, data, etc.. In summary, a lack of differences between natural and Lombard speech could be explained by a number of inherent problems. An ad hoc analysis of the results could bring more light, if no changes over time can be found. Alternatively, a deeper analysis of the RaLoCo, not warranted by the goal or scope of this study, could possibly remedy some of these.

This thesis asks participants with no previous vocal training to speak short sentences in both Lombard and normal speech. Participants speak for a total time of less than ten minutes, normal and Lombard speech combined. Summarizing the results of the previously mentioned literature, it is most likely that this thesis will find results similar to the first hypothesis: that the distance between speech parameters in the Lombard condition and the natural condition will decrease. Moon and Lindblom's (2003) cannot give us any conclusive predictions on the effect of oxygen consumption on vocal effort in Lombard speech. As Lombard speech has a higher intensity but a lower articulation rate than natural speech, it is unclear what the effects of total oxygen consumption and vocal strain will be, compared to natural speech. Participants could become

fatigued, or total oxygen consumption could be nullified by the antagonistic effects of intensity and articulation rate. The Gelfer et al. (1991) study on loud speech supports the decrease-hypothesis, as it suggests that vocally untrained participants will suffer from vocal strain after an hour of speaking. This thesis will indirectly attempt to replicate Gelfer et al. study on Lombard speech, albeit subjecting participants to an 80% shorter experiment time. Lee and Baese-Berk (2020) support two hypotheses: the no-change hypothesis, and the decrease-hypothesis. The moment that participants actively use clear speech, their intelligibility remains stable. This could potentially be the case in this thesis's study also, for the short sentences spoken in Lombard speech. However, as Lee and Baese-Berk's participants lose their motivation for 'speaking clearly', their intelligibility decreases. It is possible that this study's participants, who only have one instance in which they are asked to speak clearly, will likewise lose their motivation to speak clearly over the course of the experiment. That would result in a decrease of Lombard speech parameters. At this point in time, evidence from existing literature points towards the first hypothesis – and in so doing, implies that this thesis will find a decrease in distance of acoustic-phonological parameters between Lombard and natural speech, resulting from fatigue. Alternatively, distance between parameters can remain stable between natural and Lombard speech, at which point further investigation of individual data trends will be necessary to understand the results. If the results of this thesis show an increase in distance of parameters between Lombard and natural speech, this will likely be because of the difference in spoken time between previous studies and the current one.

In this thesis, I seek to answer the question whether Lombard speech techniques change over time spoken. The time that participants use speech, is represented by trial number, as participants use more speech as they progress on trials. Examining the effects of spoken time on Lombard speech will help us understand the phenomenon more fully. Using this thesis's results and interpretation, we can gain predictive power over many situations that involve speech production, including in noisy environments. To this end, I will analyze several acoustic parameters of natural and Lombard speech fragments. These parameters will include F0, articulation rate, and sentence spectral tilt. Also included are duration, F1, and F2 of specific vowels. Whereas sentence F0 and articulation rate are measured in relation to intelligibility, spectral tilt is a measure for 'shouted speech'. Measuring the F1 and F2 of vowels will provide an insight into the type of articulation of participants. Together, the analysis of these parameters

over trials will paint a detailed picture of the differences between Lombard and natural speech. At least three generalized outcomes resulting from this study can be hypothesized, focusing on the parameter of distance between Lombard and natural speech parameters: a decrease of distance between Lombard and natural speech over trials could indicate increasing speaker fatigue, resulting participants producing less and less pronounced Lombard speech and thus decreasing the difference between Lombard and natural speech; an increase in distance could imply a training phenomenon on the part of speakers, resulting in participants increasing their control over Lombard speech parameters, making their Lombard speech increasingly pronounced; lastly, seeing no change in distance over trials leaves much room for discussion, as the result does not lend itself to easy interpretation. I hope that this thesis's results will add to the understanding of the Lombard effect, and that it will encourage research to include trial effects for the understanding of acoustic phenomena.

2. Methods

2.1. Corpus research

The data used for this study is taken from the RaLoCo, a dataset containing speech from 78 native Dutch speakers (Shen & Janse, 2019). Speakers were between the ages of 18 and 32, and none reported hearing or speech pathology. The dataset consisted of 48 sentences per speaker. Each of these sentences was produced under both Lombard and natural conditions. Each sentence consisted of 12-16 syllables, and half of them contained a keyword noun with the Dutch corner vowels /i:/, /a:/, or /u:/ (see table 1).

1	With target noun /a:/	Peter heeft de **zaak** toevallig gisteren opgelost. *Peter happened to solve the **case** yesterday.*
2	No target noun	Laura heeft het bed naast de kledingkast geplaatst. *Laura put the bed next to the wardrobe.*

Table 1: Two example sentences that were read by the participants. Sentence 1 contains a keyword, sentence 2 is a filler sentence that does not contain a keyword.

The sentences were presented to the participants through one of four randomized lists on PowerPoint slides. Under natural conditions, speakers were instructed to read out the sentences fluently and naturally. For the Lombard condition, participants heard speech-shape noise played through headphones at 82 dB SPL. Participants were instructed to read out sentences as clearly as possible while under this condition.

2.2. Acoustic Analyses

The goal of was study is to shed light on whether Lombard speech parameters change as a function of time spoken. As a means to achieving this purpose, two acoustic analyses of trial effects were conducted. Those analyses provide insight into whether participants increase or decrease speech alterations towards the end of the experiment relative to the beginning of the experiment, or whether their Lombard techniques remain stable. In the first analysis, acoustic parameters as described in the following paragraph were measured over the sentences as a whole. In the second analysis, acoustic parameters of target vowels were analyzed. The dependent variables in the analyses were the acoustic parameters that will be measured. For the

sentence analysis, these weree different parameters than for the vowel analysis. The remaining variables wer identical for both analyses: the independent variable is the speech condition (Lombard or natural), sentence number was a fixed variable, and speaker sex was a random variable.

2.2.1. Sentence analysis

Sentence analysis was carried out over all items in the corpus. With 78 participants, each of whom spoke 24 sentences under both natural and Lombard conditions, this accounts for 7.488 entries in total. As the first acoustic analysis was done over whole sentences, intervals of silence had to be eliminated from recordings. These intervals would have influenced acoustic measurements and thus skewed results. For efficiency purposes, all sentences were concatenated per speaker and condition, and instances of silence >=200ms were selected and removed through PRAAT. Silence intervals were checked manually before deletion to ensure the accuracy of silence boundaries and to check whether all silences had been selected. The concatenated sentences with the silences removed were then cut back up into individual sentences, by marking the sentence boundaries. The sentences that emerge from these edits are used as the input for the first analysis.

The acoustic phenomena measured in the sentence analysis were: articulation rate, mean F0, and spectral tilt. The three measurements were automatically retrieved for each of the 7488 sentences with no sex-specific settings, and written out to a matrix. Note that by excluding sex-specific settings, the measurements had a higher chance of including errors, at least in F0 measurements. Articulation rate was calculated by dividing the number of syllables per sentence, divided by the duration per said sentence. This higher marge of error was unavoidable, as including sex-specific settings would have been too complex and time consuming for the current thesis. Dividing the articulation rate by numbers of syllables per sentence, resulted in a measure of syllables per second. The mean F0 (in Hz) of a sentence was taken as the F0 parameter. Spectral tilt was calculated as a difference in maximum energy in dB between a high and a low frequency band, in order to calculate the angle of the slope. The energy from the high-frequency (HF) band (2000-5000 Hz) was subtracted from the energy of the low-frequency (LF) band (0-2000 Hz). A lower number here signifies a flatter spectral tilt, as there is less difference in intensity between the two bands. These acoustic phenomena were calculated over all individual sentences through PRAAT. Written out to a table, this provided a data matrix that shows all

acoustic information per speaker, condition, and sentence, and identifies the trial it was (see fig. 2 for an extract from the matrix).

Speaker	Type	Sentence	Trial	Duration	meanF0	SpecTilt
1	Nat	F10	29	3008	227.03	14.34
1	Lom	F10	11	3272	307.98	8.03
1	Nat	F11	31	2817	227.20	15.17
1	Lom	F11	2	3150	266.01	8.28
1	Nat	F12	40	3553	218.08	13.37
1	Lom	F12	47	4220	295.48	7.94

Table 2. Extract from data matrix, with six sentences in both natural and Lombard condition by speaker 1. These filler sentences did not contain a target word – this is of no importance to the first analysis.

In order to better identify the relationship between the three acoustic parameters across speaking styles, the correlation between articulation rate, mean F0, and spectral tilt was calculated in R using Pearson's *r*. This was done before applying the linear mixed-effects model (lme).

Lme's were applied in R to compare the effect of trial number on articulation rate, mean F0, and spectral tilt in the sentences between the two conditions. For each individual parameter, an lme was performed. Articulation rate, mean F0, and spectral tilt were entered as dependent variables into their individual models; participants and sentence number were entered as random effects; and condition and trial were fixed effects. Natural speaking was the standard intercept. The main focus of analysis was the interaction between condition and trial, as this thesis wants to investigate how speech changes as it is being used (trial) differently for natural and Lombard speech (condition).

2.2.2. Vowel analysis

The vowel analysis set out to measure articulation differences between natural and Lombard conditions. It concerned the vowels from the aforementioned keywords: /a:/, /i:/, and /u:/, as these are the corner vowels of the Dutch language – and therefore the extreme ends of vowels in the language. This analysis was carried out over the speech data of 20 participants (four men and sixteen women), providing 480 data entries in total. More was not feasible due to time

restrictions. The goal of this analysis was to identify how people alter vowel formants over time, both under natural and Lombard conditions, and how the alteration differs between these conditions. Half of the sentences spoken by participants in the RaLoCo contain the keywords with one of the three Dutch corner vowels. In order to prepare individual vowels so they could be analyzed, these vowels were manually selected in PRAAT. As this analysis contains less data than sentence analysis in both speaker amount and sentence per condition, I have chosen a rougher split for the calculation of the trial effects: I dubbed the first half of spoken sentences as "early", and the second half "late". Dividing the data in two categories is more likely to catch a significant change, as there is a larger step between two large categories than between many individual ones.

Vowel duration, F1, and F2 were extracted from the previously mentioned vowels. All data was calculated through PRAAT. Settings in PRAAT needed to capture the speaking range (Hz) of all speakers, and show us all relevant formant information. Therefore, maximum formant setting was set to '5500 Hz', and the default speaker setting was set to 'men and women'. Maximum number of formants was set to 5, which is standard in most speech analyses. F1 and F2 were used in the eventual calculation. Vowel duration was calculated by subtracting the vowel's starting point from its end point, multiplied by 1000 to produce the duration in milliseconds. The vowel's midpoint, which was used for the F1 and the F2, was calculated by dividing the starting point plus the end point of the vowel by 2. At this midpoint, F1 and F2 in Hz were measured by PRAAT. By way of this process, I was able to compare data from early sentences to late sentences in both the natural as the Lombard condition.

Three linear mixed-effects models were applied to compare the effect of early and late trial type on the duration, F1, and F2 on the three vowels between the two conditions. Duration, F1, and F2 were entered as dependent variables into their individual models, and participants and sentence number were entered as random effects. Condition, vowel and trial type were fixed effects. At the standard intercept were the natural speaking condition, the vowel /a:/, and the early sentences. The main interest of this analysis was the interaction between condition and trial type.

3. Results

In this thesis, I performed two analyses as a means to answering the research question: whether speaking techniques used in Lombard and natural speech change over time spoken, and whether they do so differently under natural conditions than under Lombard conditions. In the sentence analysis, acoustic measures were aggregated over the sentences as a whole, with 7.488 data entries. Vowel analysis was carried out over keywords in sentences of 20 participants, with a total of 480 data entries. The keywords contained Dutch corner vowels /u:/, /i:/, and /a:/. The analyses carried out for the trial effects were fit to the size of the data used per analysis. Lme's were used to calculate changes in multiple dependent variables over time, to see whether natural speech patterns differ from Lombard speech patterns. In accordance with standard practice in Lombard effect studies, individual speaker and sentence variation were accounted for by adding them as random variables. The main subject of analysis was the interaction between condition and trial (type), the conditions being Lombard and natural, and trial (type) signifying the time participants used speech. In the vowel analysis, vowel type was an additional fixed effect.

3.1. Sentence-Level Measures

Sentence measures concerned the sentence as a whole. The dependent variables in this analysis were: articulation rate, mean F0, and spectral tilt, as discussed in the methods section. In this section, for all three variables first the condition effect will be discussed, followed by the interaction between condition and trial effects.

	Natural condition	**Lombard condition**
Articulation rate	5.56 syl/sec	4.71 syl/sec
Mean F0	194.39 Hz	221.30 Hz
Spectral tilt	20.03 dB	13.73 dB

Table 3: Average of the acoustic measures in the two conditions

Before calculating the lme model between condition (natural/Lombard) and trial (1-48), I calculated the underlying correlations in order to clarify the relationship between the dependent variables (articulation rate, mean F0, and spectral tilt).

3.1.1. Correlations

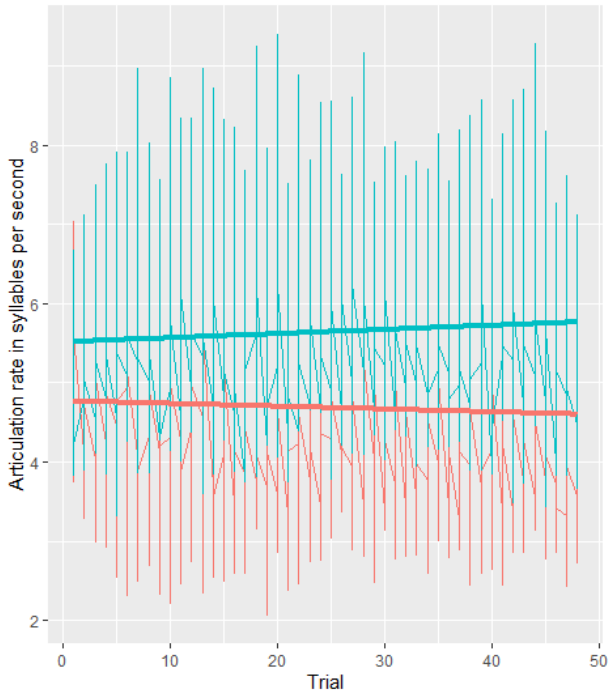
Over all 7488 data entries, multiple Pearson's correlations were used to calculate the interaction between mean F0, articulation rate, and spectral tilt, to analyze the linear relationship between these variables. The associations between the three parameters were all significant ($p < 0.001$). The negative correlation between mean F0 and spectral tilt ($r = -.38$) indicates a decrease of spectral tilt as mean F0 increases and vice versa, meaning that higher pitched speech goes together with a more evenly distributed frequency intensity. The negative interaction between mean F0 and articulation rate ($r = -0.44$) indicates a decrease of articulation rate as mean F0 increases and vice versa. Higher pitched speech thus coincides with a slower rate of speech. The positive correlation between spectral tilt and speech rate ($r = 0.46$) indicates that either variable increases when the other increases: faster speech is accompanied by a steeper spectral tilt.

3.1.2. Articulation Rate

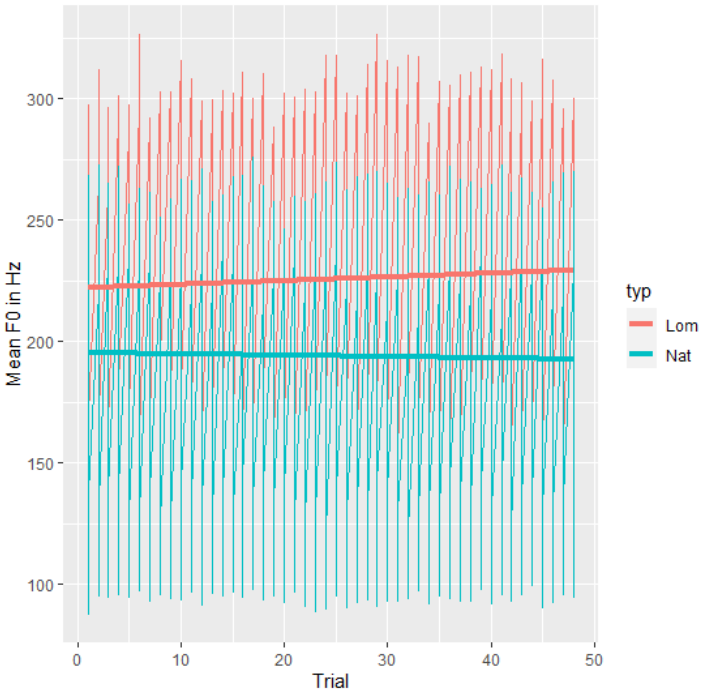
The condition effect on articulation rate in syllables per second was significant ($b = -0.85$, s.e.$= 0.02$, $p < 0.001$), with a higher articulation rate in the natural condition than in the Lombard condition. With the natural condition at the intercept, the trial effect was significant ($b = 0.003$ s.e.$= 0.001$, $p < 0.001$), indicating that articulation rate increased over trials. As can be seen in figure 1.A, the interaction between the condition and the trial effect was significant ($b = -0.004$, s.e.$= 0.001$, $p < 0.001$). This suggests that the pattern over trials differed between the two speaking styles. After releveling the intercept to the Lombard condition, the trial effect was no longer significant ($b = 0{,}000$, s.e.$= 0{,}000$, $p > 0.05$), confirming the interaction between condition and trial effects: in the natural condition, articulation rate increased significantly, while there was no significant change in articulation rate in the Lombard condition.

3.1.3. Mean F0

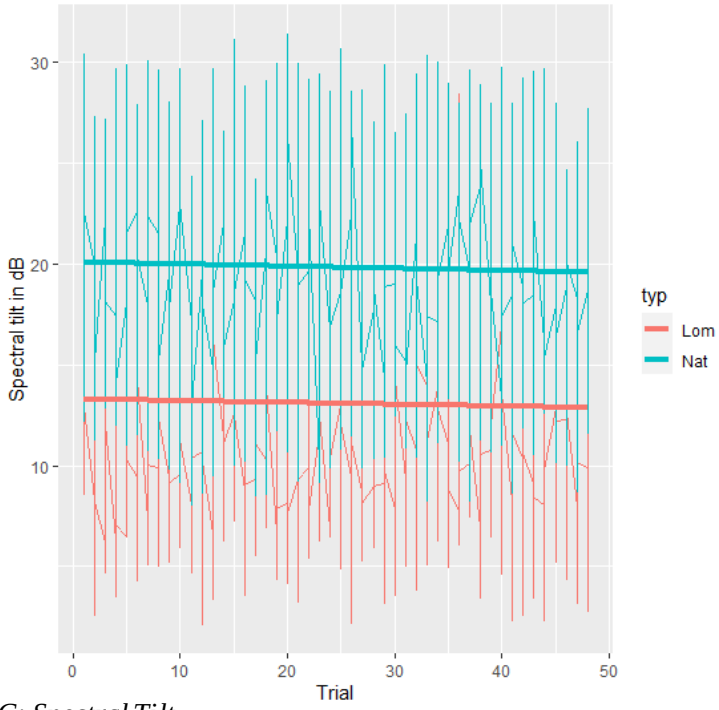
The condition effect on mean F0 in Hz was significant ($b = -27.91$, s.e. $= 0.60$, $p < 0.001$), with a lower mean F0 in the natural condition than in the Lombard condition. With the natural condition at the intercept, trial effect was not significant ($b = -0.02$, s.e. $= 0.02$, $p > 0.05$). As can be seen in figure 1.B, the interaction between condition and trial effect was significant ($b = 0.19$, s.e. $= 0.02$, $p < 0.001$), suggesting that the pattern over trials differed between the two speaking styles. Indeed, with the Lombard condition at the intercept, mean F0 significantly increased over trials ($b = 0.18$, s.e. $= 0.02$, $p < 0.001$).



A: Articulation rate



B: Mean F0



C: Spectral Tilt

Fig 1: The alteration of articulation rate, mean F0 and spectral tilt over trials, per defined condition, with individual points plotted in the background, and a bolded trend line for clarification.

3.1.4. Spectral tilt

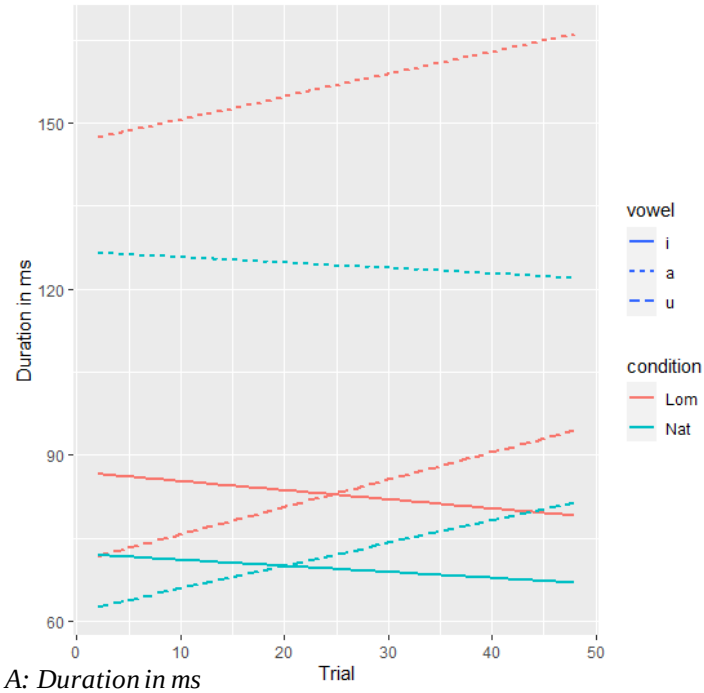
The condition effect on spectral tilt in volume (in dB) was significant ($b = -6.30$, s.e. = 0.16, $p < 0.001$), with a higher spectral tilt value in the natural condition than in the Lombard condition, signifying a greater difference between high and low frequencies in speech produced in the natural than in the Lombard condition. The interaction between condition and trial effect was significant ($b = 0.02$, s.e. = 0.006, $p < 0.005$). With the natural condition at the intercept, spectral tilt did not significantly change over trials ($b = 0.00$, s.e. = 0.00, $p > 0.05$). With the Lombard condition at the intercept, there is a decrease of spectral tilt over trials ($b = -0.03$, s.e. = 0.005, $p < 0.001$). Figure 1.C visualizes the change of spectral tilt over time. It suggests that in both conditions there is a trend of decreasing spectral tilt over time (although this trend is not significant for the natural condition), but that this decrease is significantly larger in the Lombard condition. As spectral tilt is an index of yelling, these measurements show that participant's speech becomes less 'shouty' over time.

3.2. Vowel Analysis

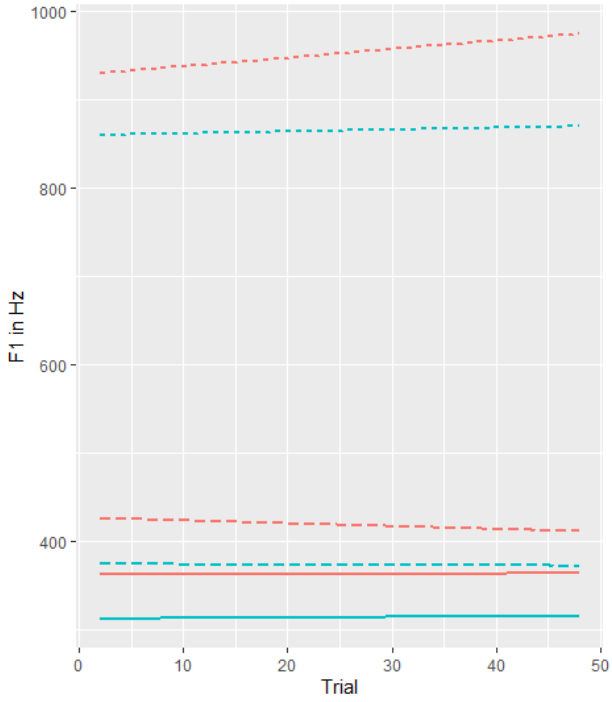
The vowel analysis was conducted to see if there are articulation differences between Lombard and natural speech. The vowel measure concerned the acoustic information of the Dutch corner vowels /a:/, /i:/, and /u:/, which were implemented in keywords in half of the sentences in the corpus. Per individual variable I have calculated the effects of vowel (/a:/, /i:/, and /u:/), condition (Lom and Nat), and trial type (early and late), as well as their mutual interactions. These variables have been explained in the method section. All lme models were conducted with natural speech, the vowel /:a/ and 'early' trial mapped on the intercept.

	Natural condition	Lombard condition
Vowel /a:/ duration	124.53 ms	155.19 ms
Vowel /a:/ F1	873.17 Hz	953.27 Hz
Vowel /a:/ F2	1535.91 Hz	1575.02 Hz
Vowel /i:/ duration	70.45 ms	83.27 ms
Vowel /i:/ F1	312.52 Hz	363.35 Hz
Vowel /i:/ F2	2488.17 Hz	2459.92 Hz
Vowel /u:/ duration	71.53 ms	85.24 ms
Vowel /u:/ F1	374.55 Hz	432.819 Hz
Vowel /u:/ F2	1054.22	1242.77 Hz

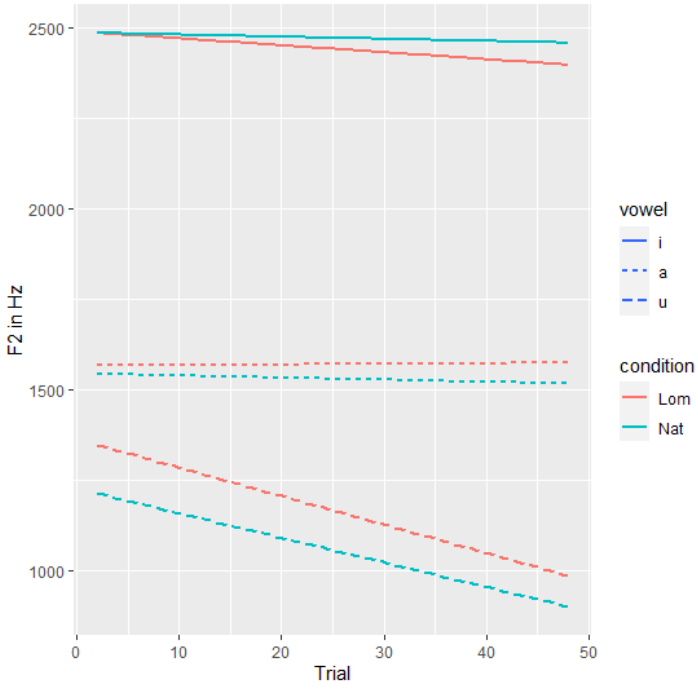
Table 4: Average acoustic measures of the vowels in the two conditions



A: Duration in ms



B: F1 in Hz



C: F2 in Hz

Fig 2: Trend lines of the alteration of duration, F1 and F2 in vowels over trials, per vowel, and per condition.

3.2.1. Duration

With the early position at the intercept, the condition effect on duration (in milliseconds) was significant for the vowel /a:/ ($b = 30.66$, s.e. = 3.09, $p < 0.001$), for vowel /i:/ ($b = 12.82$, s.e. = 2.77, $p < 0.001$), and vowel /u:/ ($b = 3.71$, s.e. = 4.07, $p < 0.001$), with a shorter vowel duration in the natural condition than in the Lombard condition for all three vowels. Interactions between vowel and condition were significant: the vowel /a:/ showed a larger condition difference than the vowels /i:/ ($b = -17.84$, s.e. = 5.11, $p < 0.001$) and /u:/ ($b = -16.00$, s.e. = 4.15, $p < 0.001$), meaning that the duration of the vowel /a:/ changed more over conditions, than the other two vowels. Thus, the increased duration of the vowel /a:/ in Lombard speech, was significantly longer than the other two vowels.

With the vowel /a:/ at the intercept, trial type was not significant ($b = -2.629$, s.e. = 4.065, $p > 0.05$). The interaction between vowel and trial type was not significant either ($b = 5.694$, s.e. = 6.81, $p > 0.05$), indicating that trial type was also not significant for any of the other vowels. Additionally, there was no significant interaction between condition, vowel, and trial type ($b = -5.134$, s.e. = 6.97, $p > 0.05$), showing that none of the other vowels had a significant change in duration over trials between the conditions. Figure 2.A visualizes the trends of change in duration in both Lombard and natural condition for all vowels.

3.2.2. F1

With the early position at the intercept, the condition effect on F1 in Hz was significant for the vowel /a:/ ($b = 80.10$, s.e. = 13.17, $p < 0.001$), for vowel /i:/ ($b = 50.83$, s.e. = 11.75, $p < 0.001$), and vowel /u:/ ($b = 58.27$, s.e. = 17.36, $p < 0.001$). As can be seen in figure 2.B, the F1 of all vowels is consistently higher in the Lombard condition than in the natural condition. This change occurred in equal intensity for each vowel.

Trial type was not significant for neither vowel /a:/ ($b = -22.71$, s.e. = 15.90, $p > 0.05$), vowel /i:/ ($b = 6.14$, s.e. = 18.33, $p > 0.05$), nor vowel /u:/ ($b = -1.37$, s.e. = 17.64, $p > 0.05$). There was also no significant interaction between condition and trial type for vowel /a:/ ($b = 22.31$, s.e. = 20.77, $p > 0.05$), /i:/ ($b = -5.89$, s.e. = 23.66, $p > 0.05$), or /u:/ ($b = -20.35$, s.e. = 20.96, $p > 0.05$), confirming that all the vowels showed the same absence of significant trial type interaction.

3.2.3. F2

With the early position at the intercept, the condition effect on F2 in Hz was not significant for the vowel /a:/ ($b = 39.11$, s.e. = 34.19, $p > 0.05$), and the interaction of condition and vowel relative to /a:/ was significant for the vowel /u:/ ($b = 149.440$, s.e. = 56.47, $p < 0.01$) and not for the vowel /i:/ ($b = -67.359$, s.e. = 45.85, $p > 0.05$). With the vowel /u:/ at the intercept, the condition effect was significant ($b = 188.55$, s.e. = 44.95, $p < 0.001$). As can be seen in figure 2.C, F2 is lower in the natural condition than in the Lombard condition.

With the natural condition at the intercept, trial type was not significant for vowel /a:/ ($b = -9.83$, s.e. =43.06, $p > 0.05$), and the interaction between trial type and vowel was not significant ($b = -22.489$, s.e. = 67.07, $p > 0.05$). This can be seen in figure 2.C.

With the vowel /a:/ at the intercept, there was no significant interaction between condition and trial type ($b = 2.221$, s.e. = 54.65, $p > 0.05$). There was however a significant interaction with the vowel /u:/ ($b = -161.080$, s.e. = 76.91, $p < 0.05$). With /u:/ at the intercept, this three-way interaction was significant ($b = -158.86$, s.e. = 54.12, $p < 0.005$). While there was no significant effect of trial type in the natural condition, the vowel /u:/ showed a significant trial type effect in the Lombard condition ($b = -191.18$, s.e. = 57.49, $p < 0.005$), where the iterations produced in the 'late' stage had a lower F2 than in the 'early' stage. As seen in figure 2.C, this results in an F2 that decreases more in the Lombard than in the natural condition.

3.2.4. Formant changes

Figure 3 shows the F1 and F2 of all three vowels together, in both the Lombard and the natural condition. F1 increases significantly for each vowel in the Lombard condition, meaning that the position of the tongue is lower for each vowel compared to natural speech. This increase is significantly larger for the vowel /a:/ than for the vowels /u:/ and /i:/. F2 only changes significantly for the vowel /u:/, with a higher F2 in the Lombard condition. This signifies that the position of the tongue is in a more anterior position in the mouth in comparison to the natural condition, only for the vowel /u:/. This suggests a more articulated version of the vowel. The F2 of the other two vowels does not significantly change between the conditions.

While the vowel /a:/ shows an increased lower tongue position in the Lombard condition in comparison to the other two vowels, the /u:/ shows an increased anterior tongue position. These differences between the vowels suggest a more emphasized realization of the vowels in Lombard

speech, indicating clearer articulation. As there are no trial type effects, nothing can be said about the change of articulation over time spoken.

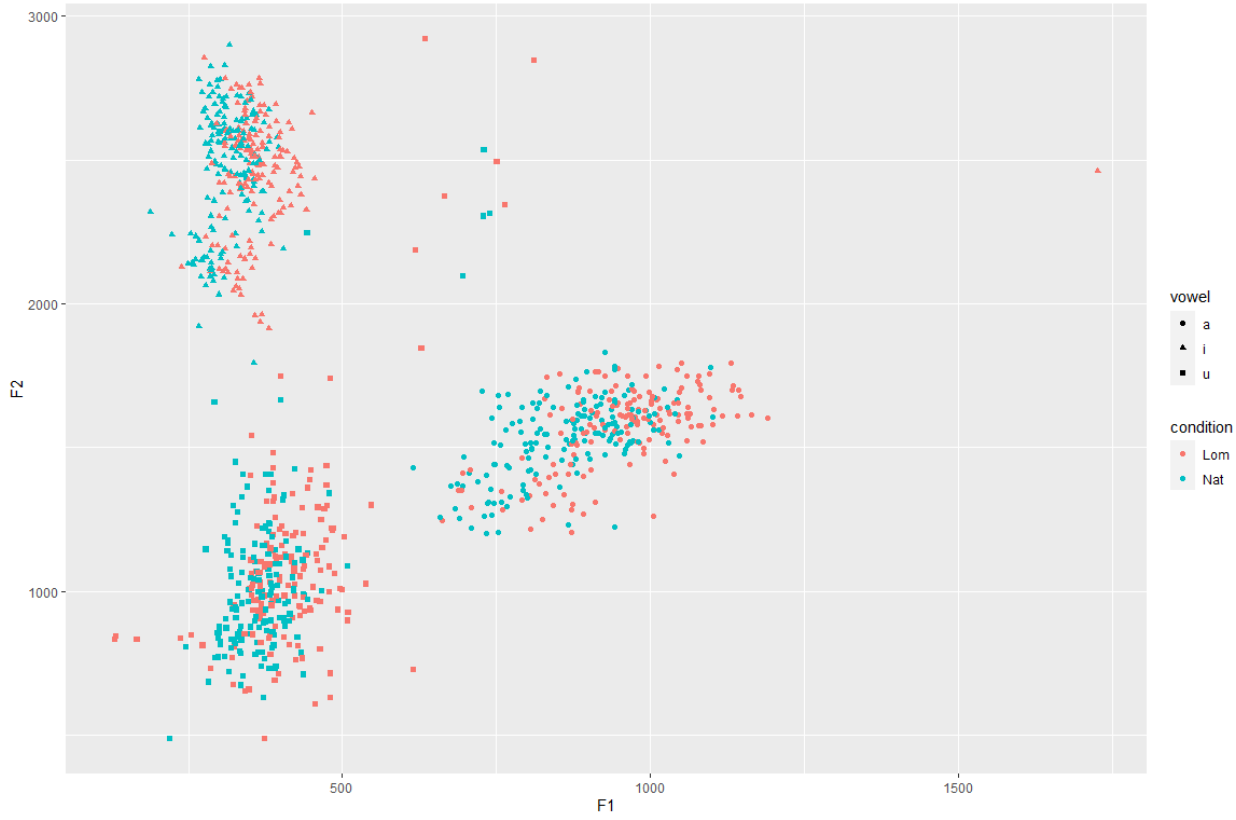


Fig. 3: Data points of the vowel triangle per condition.

4. Discussion

The purpose of this thesis was to investigate the effects of prolonged Lombard speech use on its acoustic parameter changes. To this end, the articulation rate, mean F0, and spectral tilt of sentences under natural and Lombard conditions (with instructions to "speak clearly") were measured, along with the duration, F1, and F2 of the vowels /a:/, /i:/, and /u:/. In this discussion, I will first summarize Lombard speech characteristics found in this thesis's analyses, and follow it up with a summary of noted trial effects. Lastly, I will discuss the possible implications of these results.

4.1. Condition effects

The acoustic parameter differences between Lombard and natural conditions measured in the sentence analysis, and to some extent in the vowel analysis, correspond well to the characteristics of Lombard speech as described in the literature.

In line with Junqua (1993), Lau (2008), Garnier et al. (2008), and Garnier and Henrich (2014), sentence articulation rate in this study is lower under Lombard conditions than under natural conditions. The mean F0 of these sentences is higher under the Lombard condition, in accordance with Summers, Pisoni, Bernacki, Perlow, & Stokes, 1988; Junqua, 1993 and 1996; Lau, 2008; Lu and Cook, 2009; Garnier and Henrich, 2014; Folk and Schiel, 2011. As in the literature (Summers et al., 1988; Lu & Cooke, 2009), sentence spectral tilt is flatter under Lombard conditions than under natural conditions. In further accordance with Junqua (1993), Lau (2008), Garnier et al. (2008), and Garnier and Henrich (2014), vowel duration differs significantly between Lombard and natural conditions, with increased vowel duration under the Lombard condition. Also in line with Summers et al. (1988), Junqua and Anglade (1990), and Garnier and Henrich (2012), the first formant of the studied vowels under the natural condition differs from the Lombard one, with F1 being significantly higher under the latter condition.

This study's results do not show any differences between natural and Lombard conditions in the F2 for the vowel /a:/ and /i:/, but only in the /u:/. For /u:/, F2 is significantly lower under the natural condition. Summers et al. (1988) did not find a significant condition effect for the F2 in vowels either. However, Summers et al. did find a pattern in the interaction between word type and environment condition: utterances containing high F2 frequencies showed a decrease in F2 under noisy conditions, whereas words containing low F2 frequencies showed an increase under

noisy conditions. Looking at the visualization of F2 changes found in in figure 2.C of this study, there appears to be an indication of replication of Summer et al.'s study: the F2 of high F2 frequency-vowels tends to decrease in the Lombard condition, and the F2 of low F2 frequency-vowels tends to increase. Although the differences do not reach the level of significance, the lines of vowels /a:/ and /u:/, which have a relatively low F2 frequency, suggest a higher F2 in the Lombard condition. The vowel /i:/, which has a higher F2 frequency, shows a (non-significant) tendency toward lower F2 under the Lombard condition, but only in later trials. These patterns thus suggest that Summers et al.'s results could perhaps be confirmed if more data was used for analysis.

As a result of this study's significant findings, we can conclude that speakers articulate more clearly under Lombard conditions than they do under natural conditions. F1 in vowels increases under the Lombard condition, indicating a significantly lower tongue position, especially in the vowel /a:/. F2 increases significantly in the vowel /u:/, indicating a more anterior tongue position than under normal conditions. These shifts and differences in vowel behavior create a greater contrast between the vowels, suggesting a more pronounced articulation in the Lombard condition in comparison to the natural condition.

Acoustic adaptions of the vowels found by this thesis are in line with the literature. Where the changes do not reach significance, a tendency of shift of formant frequencies, in line with the literature can be seen. It is therefore likely that the dataset used for vowel calculation was not large enough to capture the F2 change for the vowels /a:/ and /i:/. The assumption that this shift would in fact occur, can easily be assessed by use of a larger data set, which is already available in the RaLoCo. As discussed in the introduction, the scope of this thesis precludes the use of the entirety of the dataset, however. In conclusion, almost all predicted acoustic changes occur in the anticipated direction. This shows that the RaLoCo data used for this thesis has successfully captured the Lombard phenomenon, and that the trial analysis discussed below is grounded correctly.

4.2. Trial effects

This thesis's main question is whether the difference between speech under Lombard and natural conditions would change as speakers use Lombard speech over extended periods of time. As this change was calculated with reference to a possible acoustic shift over trials, it was dubbed 'the trial effect'. Three possible outcomes were discussed in the introduction. The

difference between Lombard and natural speech variables could decrease (possibly suggesting fatigue after using Lombard speech), increase (possibly suggesting a training phenomenon after using Lombard speech), or show no change over time.

Sentence analysis results show that, as participants speak for a longer time, all acoustic parameters under the Lombard condition become further distanced from the natural condition baseline. The differences in articulation rate, mean F0, and sentence spectral tilt between Lombard and natural conditions increase over trials. Lombard speech becomes increasingly slower, higher pitched, and more 'shouty' than normal speech. The vowel analysis, in contrast to the sentence analysis, had results that were hard to interpret. F1 showed no significant trial effects in vowel analysis. The vowel /u:/ had a lower F2 in the 'late' trials than in the 'early' ones, meaning its F2 was lower in the second half of the trials than in the first half. In contrast, none of the other vowels showed a significant difference between early and late trials. The vowel /u:/ was also the only vowel showing a significant difference in F2 between Lombard and natural conditions, in the sentence analysis. The difference between conditions thus became larger for /u:/ than it did for the other two vowels. However, F2 differences between Lombard and natural conditions seem to decrease from early to late trials, in contrast to the pattern of the sentence analysis. This discrepancy has multiple explanations: the vowel /u:/ may behave differently in Lombard conditions than most other components of the sentences, but the discrepancy could also simply be due to inconsistent measurement. Sentence and vowel analysis did not have identical data processing methods: the data used for the vowel analysis was smaller than the data used for the sentence analysis, and each individual data entry was much shorter. Because of the difference in the two analyses' results, and the different data processing methods that were used, I will first analyze the vowel analysis' peculiarities, before moving on to a more general discussion of the results.

4.2.1. Vowel analysis

A majority of vowel analysis results did not return significant results. Findings of the vowel analysis do not replicate the findings of previous literature. Strikingly however, sentence analysis perfectly replicated the results of earlier studies. Therefore, it is unlikely that the RaLoCo has failed to capture the Lombard effect. Rather, the vowel analysis used in this thesis is likely flawed, in the sense that its sample was not large enough to support the method of analysis used. The relatively limited dataset combined with the different method used in the analysis, have

made the vowel analysis results an unreliable source for speculation about the Lombard effect. Although this does not invalidate this entire thesis, vowel analysis results on their own cannot generate many robust contributions to the discussion of Lombard speech.

The results of the vowel analysis and the low rate of significance, the insufficiently precise calculation on small units of data, and the use of a different calculation method in the vowel analysis as compared to the sentence analysis make the vowel analysis unsuitable for speculations about the topic of this thesis. The vowel /u:/ has a higher F2 under the Lombard condition than under the natural condition. Additionally, as seen in figure 2.C, the vowel's frequency falls more acutely under the Lombard condition than under the natural condition. Under the Lombard condition, the vowel /u:/ has a lower F2 in the 'late' trials than in the 'early' trials, while there is no significant difference of F2 between early and late iterations under the natural condition. In contrast to sentence analysis results, the distance between the Lombard and natural conditions seems to *decrease* over trials. Furthermore, unlike in the sentence analysis, the other trial effects in vowel analysis are non-significant. This is a messy combination of results, which brings up doubts about the methodology. Indeed, the sound-instances taken from the vowel, that were used in the analysis, were short (<200 ms), and speaker-sex-based adjustments were excluded from the measurements. Both of these properties in the analysis lead to a larger margin of error. Taking into account these arguments, inaccuracies in measurement are more likely to occur during vowel analysis than sentence analysis. In order to eliminate measuring inaccuracies, a larger data pool should have been used for vowel analysis. This could have produced clearer results for this study. However, time restrictions limited this study to include not more than 20 participants, making the analysis unreliable. Additionally and as previously stated, the method used for calculating trial effects in the vowel analysis was different than the method for sentence analysis. Using two different calculation methods was a conscious and deliberate choice, because the calculation concerned two different units, and changes from trial to trial were smaller than in the sentence analysis, for which the previous calculation method was insufficiently sensitive. However, using two different calculation methods allows for discrepancies between the results. In combination with the unreliably small dataset, this makes the vowel analysis unsuitable for answering my research question.

In summary, I believe the dataset used to measure changes in the vowel parameters was too small for the object of this analysis. Furthermore, the calculation method used, led to additional

risk being incurred. As a result, the results of this study's vowel analysis are insufficiently clean to provide fertile ground for speculation about the prolonged use of Lombard speech. As such, I will continue my efforts with added focus on the sentence analysis results, which have proven substantial, clear and useful.

4.2.2. Implications

In the introduction, three hypotheses were posed: the data could show a decreased distance between Lombard speech parameters and natural speech parameters over time, suggesting speaker fatigue; the data could show an increased distance between Lombard and natural speech parameters over time, suggesting speech training; and the data could show no consistent change over time between the two conditions. As noted in the previous paragraph, for all acoustic parameters in the sentence analysis, the distance between the natural and the Lombard conditions increased over trials. This result is largely in accordance with the second hypothesis, suggesting a training effect. Before I engage with these findings and explore their implications, I will first discuss the two hypotheses that did not match with the results. Subsequently, I will engage in discussion and elaborate on possible implications of the training hypothesis.

Initially, my motivation to write this thesis was the question of whether Lombard speech induced fatigue in people's speech. After all: Lombard speech is characterized by a higher intensity and an sonically challenging environment for speakers. Both of these characteristics have been shown to induce fatigue in speakers (Rotton, Olszewski, Charleton & Soler, 1978; Gelfer, Andrews, & Schmidt, 1991; Kristiansen et al., 2014). In this study's results, the presence of fatigue would be represented by a decrease in defined 'Lombard speech parameters' (as discussed in paragraph 1.1) under Lombard conditions. If that decrease was visible over time, one would have found evidence for the first proposed hypothesis, and reason for further inquiry. The results of this thesis do not reflect this prediction: there was no systematic decrease in relevant parameters. Why the aforementioned studies find indications of speaker fatigue while this thesis does not, is not clear. It may be as a result of differences in the tasks that speakers had to perform: participants in the Gelfer, et al. (1991) and the Kristiansen et al. (2014) studies had to speak for a significantly longer time than the participants in the current thesis; and the Rotton, et al. (1978) study did not measure speech production, but performance on facial recognition. It is possible that fatigue only occurs after a longer period of time spoken. In the Gelfer et al. (1991) study, participants were required to read for one hour before their final speech parameters were

measured. The latter's untrained speakers showed clear signs of fatigue after the one hour period. In a study by Kristiansen et al. (2014), subjects were tested after a full day of teaching. In this thesis's data, speakers spoke 48 sentences with a maximum of 16 syllables twice (once per condition), which would not take up more than 10 minutes in total. Likely, fatigue in Lombard speech is a function of time spoken, but only after considerable time: the assumption being that if participants are required to speak in a noisy environment for a longer period of time, it will be more apparent. The only study discussed in this thesis, that organized its temporal factors similarly to this thesis is Rotton et al. (1978). However, Rotton et al. do not study speech production directly, and facial differentiation and speech production are processes too different to consider comparable for the purposes of my results. It is therefore unsurprising that this thesis's results show no decrease in acoustic parameters, and thus deviate from the first hypothesis. As an interesting side note, Rotton et al. did find a difference in background noise type on task performance: meaningful speech interfered with the tasks performed in their study, more so than unintelligible background noise. With regards to this thesis's data, speech shaped was used to elicit Lombard speech. Perhaps if a type of background noise with a larger interference effect had been used, it would have produced a more powerful fatiguing effect, in turn making fatigue measurable over a shorter time span. In conclusion, the absence of fatigue-indicating results could be due to the short timespan over which Lombard speech was used. A more challenging speech environment could possibly make the fatiguing effect appear (earlier). In future fatigue and Lombard speech studies, it is advisable to have speakers use Lombard speech for a longer period of time, take the type of background noise type into account when composing the study, and measure the speech parameters continuously in order to find the range when fatigue starts setting in.

The third hypothesis of this study predicted that there would be no measurable difference between the Lombard and natural conditions over time. If this had been the thesis's outcome, it would have been difficult to draw insightful conclusions from it, as many potential causes can explain the absence of a parameter difference between Lombard and natural conditions. However, one of the implications that follow from the invalidity of the third hypothesis, is that there is some evidence in the reflex-versus-act debate. One of the properties of a reflex is that it does not change over time, as it is an automatic response. The results of this study however show a clear change of speech parameters under both natural and Lombard conditions, and clear

differences between the conditions. This might imply that what was measured, is not (solely) a reflex. Before taking this as evidence for (or against) the reflexive nature of Lombard speech however, we need to keep into account that participants were instructed to "speak as clear as possible". It is possible that this thesis's data and results are skewed by the clear speech phenomenon. As noted extensively in the introduction, speakers are often instructed to "speak clearly" in Lombard speech research, mixing the two speech phenomena. They are so similar as to be identical in some respects. But for the purposes of the reflex-versus-act debate, the 'clear speech' instructions given to speakers tend to conflate results which can otherwise be differentiated. To reach a definite conclusion about the reflexive nature of Lombard speech, a study should be constructed with a built-in safeguard to control for participants' conscious use of clear speech.

The remaining hypothesis is the training effect one, in which speech parameters in Lombard and natural conditions increasingly change into different directions as more speech is used. The results of this study indeed confirm these exact predictions, in turn seemingly confirming the training hypothesis. Speakers showing increasingly distinguished features of Lombard speech as they speak more, could be explained by a training phenomenon, or by their adapting to the challenging environment. If interpreting the results as a training effect, it follows that speakers' improvement would be predicated on their continuous and improving use of Lombard speech. The trained singers in the Gelfer et al. study (1991) showed signs of vocal improvement over time, represented by the decrease of vocal qualities that damage the vocal organs, such as jitter (see their methods section for further information). That finding contrasts with the results of untrained singers in their study, whose vocal parameters showed clear signs of fatigue after an hour of loud reading. Singers have been trained to avoid fatigue-inducing speech techniques. This is possibly what protected them from experiencing and incurring fatigue during the testing period, and allowed them to continue improving their vocal techniques. The subjects of this thesis, in that case, should be more comparable to the untrained speakers in the Gelfer et al. study. The speakers in this thesis have also not received any prior vocal training, nor were they selected for those qualities. This thesis's results however contrast with those of the Gelfer et al. study: this thesis's participants show an increase in the measured parameters, the more they use the tested speech, rather than showing signs of fatigue. A possible explanation is that this thesis's participants have not yet reached the point in time where fatigue starts to set in, whereas the

untrained singers of the Gelfer et al. study have. This would be owing to the relatively short duration of this thesis's experiment (less than 10 minutes) as compared to the Gelfer et al. study (1 hour). In consideration of this difference in experiment setup, it could hold that, for a particular length of time, speakers would increase their control over speech parameters, until a critical point is reached whereupon fatigue would start to set in. A parallel can be drawn to any other activity that involves the muscles, which usually involves a warm-up period, a build-up towards peak performance, and then a cool-down where effort wanes as fatigue sets in. Gelfer et al.'s trained singers do not reach their point of fatigue, and neither do the participants in the current thesis. Their reasons are then very different: while the trained singers effect techniques meant to stave off fatigue, the untrained speakers do not show fatigue as a result of having to speak for a very short time (10 minutes). The results of this thesis, combined with those of the Gelfer et al. study, thus imply that people can improve their Lombard speech parameters for a certain duration of time, until they reach a vocal performance limit. This limit could perhaps differ per individual and even per situation. Based on this thesis's data, we can only say for certain that this limit is likely to be found somewhere beyond the time people in the RaLoCo spoke for most people, assuming speech shape noise conditions. This vocal limit can likely be delayed, or even avoided entirely with sufficient training, such as enjoyed by one group in the Gelfer et al. study. Another likely conclusion is the inverse: that untrained speakers overexert their vocal system after some time, and begin to show signs of fatigue, as evidenced by the untrained speakers in Gelfer et al.. When trained performers, who have learned to use their voice without exhausting their vocal system, partake in a long-lasting task, they can avoid the point of exhaustion, thus continuing the training effect experienced in the earlier stages. Additionally, if trained performers were to be trained in this way, they may avoid fatigue and show a continued learning curve in Lombard speech parameters. For future studies, it would be interesting to search for the time range in which vocal fatigue occurs in untrained participants, and, to research whether introducing formal vocal training can prolong this time range.

Apart from an increase in control over Lombard speech parameters, habituation to the challenging speech environment may also be an explanation for the results of this study. The noise presented to speakers is a source of interference to their cognitive functions (Rotton et al., 1978, Kristiansen et al., 2014). According to the Rottan et al. study, people's frustration tolerance decreases together with their ability to differentiate faces. Similarly, background noise

could have caused cognitive interference for speakers in this thesis's study as well. According to Namba and Kuwano (1988), some people eventually habituate to background noise. In their experiment, participants were presented with various background noises - both speech and non-speech - while performing three types of mental tasks. Participants were asked to indicate the level of noisiness with a button, and if they did not report on noise for over 30 seconds, participants were considered to be habituated to the background noise. Their results showed a clear dichotomy in the noise habituation of participants: one group of people showed clear habituation to the background noise over time, while another group continued to report noise throughout the experiment. Of course, we should keep into account that mental tasks do not directly carry over to speaking tasks. Nonetheless, it is possible that some (or most) participants in this thesis also habituated to the background noise. If background noise was causing interference at the beginning of their task, people could have grown accustomed to it over trials. This would have removed the interference effect and allowed participants to improve Lombard speech parameters in their speech. However, according to Namba and Kuwano's results, many people do not habituate to background noise. As the results of this study are so clear in their increase of Lombard parameters over trials, habituation to background noise cannot be the only explanation for my results. More likely, the previous training explanation plays a role, or a combination of both effects is responsible for the results.

To summarize the implications: this study's results show that distance between Lombard and naturally produced speech parameters increases over time, with speech parameters under Lombard conditions moving towards Lombard speech characteristics as participants use Lombard speech over a longer amount of time. Between the three hypotheses posed in the introduction, this pattern was predicted by the training hypothesis. The fact that the Lombard speech parameters change consistently over time spoken in this study suggests that the Lombard effect, as elicited during the collection of this dataset, is not (merely) a reflex. The task performed in the RaLoCo elicited improved Lombard speech over time, and was short enough to prevent fatigue. While the general effect of using Lombard speech for a prolonged time seems to be the training of the parameters involved, different groups appear to react differently. Untrained speakers could measurably overexert their vocal system after some time, while trained speakers might continue to increase the training of the acoustic parameters used in Lombard speech.

5. Conclusion

The research question of this thesis was *How do acoustic speech parameters change over time spoken, and does this change differ for Lombard speech when compared to natural speech?* To answer these questions, RaLoCo data was used to measure change in acoustic parameters between Lombard and natural speech conditions over time. The measured difference in acoustic parameters between the two conditions largely corresponds with the existing literature. The points where no significant correspondence with the literature was found, all appeared in the vowel analysis, and were likely due to the small dataset used for the measurements. Because of the relatively small dataset used, the vowel analysis did not provide adequate information, and is therefore not taken into account for results discussion. The change in acoustic parameters over time and the trial effects of the remaining data are clear and consistent. The parameters from the two conditions grow further apart over time, with Lombard condition speech parameters increasingly moving toward Lombard speech characteristics. Previous literature indicates that a prolonged use of loud or clear speech leads to vocal improvement for trained singers and vocal fatigue in untrained speakers. Taking this into account, the results of this study suggest that speakers can train their speech parameters when they use Lombard, loud, or clear speech, until fatigue sets in. In order to be more definite in these conclusions, a study should be created to separate the definitions and experiments of clear and Lombard speech, as too often these two are used interchangeably. Without this distinction, there is no attributing the results of this study to the Lombard effect specifically.

References

Adkins, D. L., Boychuk, J., Remple, M. S., & Kleim, J. (2006). Motor training induces experience-specific patterns of plasticity across motor cortex and spinal cord. *Journal of Applied Physiology, 101*, 1776-1782.

Baker, R., & Hazan, V. (2010). LUCID: a corpus of spontaneous and read clear speech in British English. *DiSS-LPSS*.

Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech and Language, 28,* 543-571. doi: 10.1016/j.csl.2013.08.003

Folk, L., & Schiel, F. (2011). The Lombard Effect in spontaneous dialog speech. INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy.

Garnier, M., Bailly, L., Dohen, M., Welby, P., and Loevenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: global effects on the utterance. INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, Pittsburgh, PA, United States.

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language, 28.2,* 580-597. doi:10.1016/j.csl.2013.07.005

Gelfer, M. P., Andrews, M. L., & Schmidt, C. P. (1991). Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers. Journal of Voice, 5(2), 158–167. doi: 10.1016/s0892-1997(05)80179-1

Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2014). Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles. *Computer Speech & Language 28.2,* 629-647. doi:10.1016/j.csl.2013.09.007

Hazan, V., & Baker, R. (2010). Does reading clearly produce the same acoustic-phonetic modifications as spontaneous speech in a clear speaking style? *DiSS-LPSS*.

Hazan, V., and Baker, R. (2011). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," *Journal of the Acoustical Society of America, 130,* 2139–2152. doi: 10.1121/1.3623753.

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, 93.1,* 510-524. doi:10.1121/1.405631

Junqua, J. C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication, 20.1-2,* 13-22. doi:10.1016/s0167-6393(96)00041-6

Junqua, J., & Anglade, Y. (1990). Acoustic and perceptual studies of Lombard speech: Application to isolated-words automatic speech recognition. *International Conference on Acoustics, Speech, and Signal Processing,* 841-844. doi:10.1109/icassp.1990.115969

Vlaj, D., & Kacic, Z. (2011). The influence of Lombard effect on speech recognition. In *Speech Technologies*. doi:10.5772/17520

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of Acoustical Society of America, 115.1,* 362-378. doi:10.1121/1.1635842

Kristiansen, J., Lund, S. P., Persson, R., Shibuya, H., Nielsen, P. M., & Scholz, M. (2014). A study of classroom acoustics and school teachers' noise exposure, voice load and speaking time during teaching, and the effects on vocal and mental fatigue development. *International Archives of Occupational and Environmental Health. 87,* 851-860. doi: 10.1007/s00420-014-0927-8

Lane, H. L., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing, 14.4,* 677-709. doi:10.1044/jshr.1404.677

Lau, P. (2008). The Lombard Effect as a Communicative Phenomenon. *UC Berkeley PhonLab Annual Report, 4.4*, 1-9.

Lee, D-Y., & Baese-Berk, M. M. (2020). The maintenance of clear speech in naturalistic environments. *Journal of Acoustical Society of America, 147,* 3702-3711. doi: 10.1121/10.0001315

Lombard, E. (1911). Le signe de l'élévation de la voix [The sign of the elevation of the voice]. (P. H. Mason, Trans.). *Annales des Maladies de L'Oreille et du Larynx, 27*, 101-109.

Lu, Y,. & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of Acoustical Society of America, 124.5* 3261-3275.

Lu, Y., & Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication, 51.12*, 1253-1291. doi: 10.1016/j.specom.2009.07.002

Moon, S.-J., & Lindblom, B. (2003). Two experiments on oxygen consumption during speech production: vocal effort and speaking tempo. In *International Congress of Phonetic Sciences, Barcelona, Spain*. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-11061

Namba, S., & Kuwano, S. (1988). Measurement of habituation to noise using the method of continuous judgment by category. *Journal of Sound and Vibration, 127(3),* 507-511. doi: 10.1016/0022-460X(88)90376-8.

Pépiot, E. (2014). Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *Speech Prosody, 7,* 305-309.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 29,* 434–446. doi: 10.1044/jshr.2904.434

Pomfrett, C. J. D. (2005). Neural reflexes. *Anaesthesia & Intensive Care Medicine, 6.5,* 145-150. doi: 10.1383/anes.6.5.145.65034

Raitio, T., Suni, A., Pohjalainen, J., Airaksinen, M., Vainio, M., & Alku, P. (2013). Analysis and synthesis of shouted speech. Proceedings of Interspeech 2012.

Rotton, J., Olszewski, D., Charleton, M., & Soler, E. (1978). Loud speech, conglomerate noise, and behavioral aftereffects. *Journal of Applied Psychology, 63(3),* 360–365. doi:10.1037/0021-9010.63.3.360

Rudner, M., & Lunner, T. (2014). Cognitive spare capacity and speech communication: A narrative overview. *BioMed research international*, 1-10. doi: 10.1155/2014/869726

Shen, C. & Janse, E. (2019). Description of Radboud Lombard Corpus_Dutch (RaLoCo_Dutch).

Stowe, L. M., & Golob, E. J. (2013). Evidence that the Lombard effect is frequency-specific in humans. *The Journal of the Acoustical Society of America, 134.1,* 640-647. doi:10.1121/1.4807645

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Perlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America, 84.3,* 917-928.

Tuomainen, O., & Hazan, V. (2018). Investigating clear speech adaptations in spontaneous speech produced in communicative settings. In M. Gósy & T. E. Gráczi (Eds.), *Challenges in Analysis and Processing of Spontaneous Speech* (9–25). Budapest, Veszprém: OOK Press.

Zollinger, S. A., & Brumm, H. (2011). The Lombard effect. *Current Biology, 21.16.* doi:10.1016/j.cub.2011.06.003