Non-native Lombard speech:

The acoustics, perception, and comprehension of English Lombard speech by Dutch natives

Katherine Marcoux

Funding Body
This research was funded by the European Union's Horizon 2020 research innovation programme under the Marie Skłodowska-Curie grant agreement No. 675324.

International Max Planck Research School (IMPRS) for Language Sciences
The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics
Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

ISBN: 978-94-92910-38-7
Cover design and lay-out by Andrey Vásquez
Printed and bound by Ipskamp Drukkers, Enschede
 The research reported in this thesis was conducted at the Centre for Language Studies of the Radboud University Nijmegen, the Netherlands.

Non-native Lombard speech:

The acoustics, perception, and comprehension of English Lombard speech by Dutch natives

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 30 juni 2022
om 11.30 uur precies

door

Katherine Pearl Marcoux
geboren op 20 juli 1992
te Rochester (Verenigde Staten)

Promotor:
Prof. dr. M.T.C. Ernestus

Copromotor:
Dr. E. Janse

Manuscriptcommissie:
Prof. dr. M. van Oostendorp
Dr. M.E. Broersma
Dr. R.J.J.H. van Son (Antoni van Leeuwenhoek)
Prof. dr. V.L. Hazan (University College London, Verenigd Koninkrijk)
Prof. dr. M. Simonet (University of Arizona, Verenigde Staten)

Non-native Lombard speech:

The acoustics, perception, and comprehension of English Lombard speech by Dutch natives

Dissertation to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Doctorate Board
to be defended in public on

Thursday, June 30, 2022
at 11.30 am

by

Katherine Pearl Marcoux
born on July 20, 1992
in Rochester (USA)

Supervisor:
Prof. dr. M.T.C. Ernestus

Co-supervisor:
Dr. E. Janse

Manuscript Committee:
Prof. dr. M. van Oostendorp
Dr. M.E. Broersma
Dr. R.J.J.H. van Son (Netherlands Cancer Institute)
Prof. dr. V.L. Hazan (University College London, United Kingdom)
Prof. dr. M. Simonet (University of Arizona, United States of America)

Table of Contents

Chapter 1:
Introduction

Train stations, restaurants, grocery stores, music festivals; all of these can be populated areas with many different sounds such as the movement of people and things, music, and human speech creating lots of noise. In such noisy environments, humans produce Lombard speech. Lombard speech is characterized by acoustic modifications that differentiate it from speech produced in quiet. Its acoustic characteristics have been extensively studied in native speakers of several languages, while only a handful of studies have examined non-native speakers producing Lombard speech. In this dissertation, I investigated the characteristics of non-native Lombard speech by first creating the Dutch English Lombard Native Non-Native (DELNN) corpus and then examining, on the basis of the corpus, non-native Lombard speech acoustics, non-native Lombard speech intelligibility, and non-native Lombard speech accentedness. Each of these three perspectives complements each other, resulting in a broad understanding of the characteristics of non-native Lombard speech.

1.1 Lombard Speech

Lombard speech (Lombard, 1911) is characterized by several acoustic modifications when compared to plain speech, speech produced in quiet. These changes include a shift in energy to higher frequencies, an increase in fundamental frequency (F0), an increase in intensity, changes in duration, and shifts in the vowel space (for a review, see Cooke, King, Garnier, & Aubanel, 2014). Table 1 includes a summary of well documented acoustic characteristics of Lombard speech and examples of corresponding research.

In earlier research, Lombard speech was referred to as a reflex (e.g., Junqua, 1993; Van Summers et al., 1988). Viewing Lombard speech as a reflex may suggest that the properties of Lombard speech are language universal. Further, referring to Lombard speech as a reflex implies that it is an automatic process which the speaker does not have control over. Zollinger and Brumm (2011) discuss that while Lombard speech elicits an involuntary response, since it can be controlled by the speaker (communicative intent produces a stronger Lombard effect: e.g., Junqua, Fincke, & Field, 1999; Lane & Tranel, 1971; Villegas, Perkins, & Wilson, 2021) it is "not a true reflex" (p. R614). Hence, it may be that Lombard speech shows subtle differences between languages.

Lombard speech production has been investigated in native speakers in many languages including English (e.g., Bosker & Cooke, 2018; Lu & Cooke, 2008; Pisoni, Bernacki, Nusbaum, & Yuchtman, 1985; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988),

Dutch (e.g., Bosker & Cooke, 2020), Spanish (e.g., Castellanos, Benedí, & Casacuberta, 1996), and French (e.g., Garnier & Henrich, 2014). These studies show similar acoustic trends in Lombard speech across languages.

Acoustic measure	Findings	Reference	Language
Shift in energy	long-term-average speech spectra: increase	Pittman & Wiley, 2001	English
	spectral tilt: decrease (words)	Van Summers et al., 1988	English
	spectral CoG: increase (utterances)	Lu & Cooke, 2009a	English
	spectral CoG: increase (phonemes)	Junqua, 1993	English
Fundamental Frequency (F0)	mean F0: increase (words)	Van Summers et al., 1988	English
	mean F0: increase (utterances)	Lu & Cooke, 2009a	English
First Formant Frequency (F1)	mean F1: increase (words)	Van Summers et al., 1988	English
	mean F1: increase (utterances)	Lu & Cooke, 2009a	English
Intensity	vocal levels (dB Sound Pressure Level): increase	Pittman & Wiley, 2001	English
	root mean square (rms) energy: increase	Lu & Cooke, 2008	English
	mean vocal intensity: increase	Garnier & Henrich, 2014	French
Duration	duration: increase (word)	Van Summers et al., 1988	English
	duration: increase (vowels), decrease (consonants)	Garnier & Henrich, 2014	French
	duration: increase (utterances)	Lu & Cooke, 2008	English
	duration: decrease (utterances)	Varadarajan & Hansen, 2006	English

Table 1: *A summary of several well documented acoustic measures and how they are modified in Lombard speech relative to plain speech, from examples (non-exhaustive) of research involving native speakers.*

The specific acoustic modifications of Lombard speech are influenced by several factors. Different noises, such as for example white noise and cocktail party noise, which differ in spectral content and whether they fluctuate in intensity over time, can subtly affect certain acoustics of the Lombard speech produced (e.g., Garnier, Bailly, Dohen, Welby, & Loevenbruck, 2006; Junqua, 1994). For instance, Junqua (1994) and Garnier and colleagues (2006) observed that different noise types elicited different durational changes. Additionally, the level of noise influences the Lombard speech, with higher levels of noise eliciting larger acoustic modifications for some measures (e.g., Dreher & O'Neill, 1957; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). For instance, Van Summers and colleagues (1988) analyzed Lombard speech produced at several different levels of noise (80, 90, and 100 dB), noting that each higher level elicited higher amplitudes, longer words (although not always significantly), and generally, lower spectral tilt.

Moreover, properties inherent to the speaker may play a role, with some research documenting different modifications of pitch and energy in Lombard speech for females and males (e.g., Junqua, 1993), and others discussing the importance of individual differences in Lombard speech production (e.g., Junqua, 1993; Pittman & Wiley, 2001; Shen, 2022). Lombard speech is additionally modulated by communicative intent, with effects being larger if there is a communicative purpose (e.g., Junqua, et al., 1999; Lane & Tranel, 1971). To give an example, Villegas and colleagues (2021) found that speakers produced louder Lombard speech for communicative tasks than for non-communicative tasks.

Research has shown that the acoustic modifications associated with Lombard speech allows it to be better understood in noise as compared to plain speech presented in noise, providing a Lombard intelligibility benefit (e.g., Dreher & O'Neill, 1957; Lu & Cooke, 2008; Pittman & Wiley, 2001). As with the acoustic modifications of Lombard speech, the Lombard intelligibility benefit can be influenced by several factors. As mentioned above, these include the noise used to elicit the Lombard speech (type and level; e.g., Lu & Cooke, 2008). In addition, the Signal to Noise Ratio (SNR) at which the intelligibility is tested (e.g., Van Summers et al., 1988) affects the size of the Lombard intelligibility benefit. Moreover, as in any experiment, the stimuli chosen can influence the results, for instance, depending on the frequency of occurrence, predictability, and reduction of the stimuli.

Compared to plain speech, Lombard speech is better able to withstand noise masking. This can be measured by the glimpsing proportion metric (GP, Cooke, 2006) and its extension, the high-energy glimpsing proportion metric (HEGPs, Tang & Cooke, 2016). GPs indicate the proportion of regions of speech where its energy is greater than that of the masking noise. HEGPs are an extension of GPs. HEGPs look at each frequency band separately and only select the regions in each band where the energy of speech is greater than the average speech-plus-noise energy. Lombard speech has increased GPs compared to plain speech (Lu

& Cooke, 2009b). Furthermore GPs – and HEGPs to an even greater extent – are correlated with human intelligibility scores (Tang & Cooke, 2016).

Lombard speech could be considered in the context of Lindblom's (1990) hyper- and hypo- theory of speech production. In this case, the background noise, which elicits Lombard speech, makes the listener's job harder, and the speaker can facilitate the communication by producing more hyper-articulated speech. Zhao and Jurafsky (2009) discuss their Lombard speech findings in the context of this and related theories.

Note that Lombard speech is not the same as clear speech. Lombard speech is produced when there is background noise, independent of whether there is a listener present, although the modifications can be modulated by the presence of listeners (e.g., Junqua et al., 1999; Lane & Tranel, 1971). Clear speech, on the other hand, is when the speaker experiences communication difficulties with the listener, whether environmental or listener oriented (Smiljanić & Bradlow, 2009). In their review of clear speech studies, Smiljanić and Bradlow summarize the modifications for clear speech, which include but are not limited to a slower speaking rate, shift in energy, and an increase in intensity. Lombard speech shares similar modifications. Further, based on their past research, Smiljanić and Bradlow suggest that clear speech may also involve language-specific modifications. This may suggest that Lombard speech also potentially shows language-specific modifications.

1.2 Non-natives

Being a non-native speaker or listener has its difficulties. Speakers experience a higher cognitive load when speaking in a non-native language compared to their native language (e.g., Kormos, 2006; Segalowitz, 2010). This extra difficulty for non-native speakers may affect how they speak in certain environments. Furthermore, when speaking in a non-native language, the native language is often influential. The influence of the native language may appear in many different aspects of speech production, including – but not limited to – prosody (e.g., van Maastricht, Krahmer, & Swerts, 2016), vowels (e.g., Burgos, Cucchiarini, van Hout, & Strik, 2013; Flege, 1987) and consonants (e.g., obstruents and their voice onset time; e.g., Flege, 1987; Simon & Leuschner, 2010). For instance, Burgos and colleagues (2013) found the production of non-native Dutch vowels to be influenced by native Spanish; mentioning issues of vowel length and height, among others.

In terms of speech perception by non-natives, the native language can also be influential (e.g., Flege & Wang, 1989; Lotz, Abramson, Gerstman, Ingemann, & Nemser, 1960; Miyawaki et al., 1975). For example, Miyawaki and colleagues showed that while Americans can distinguish between [ra] and [la] in speech stimuli, native Japanese listeners perform at chance level probably because the [r] and [l] is not a phonemic contrast in Japanese.

Considering that the native language influences both production and perception, it follows that the specific combination of speakers' and listeners' native languages could be influential for speech intelligibility. Bent and Bradlow (2003) found that high proficiency non-native speakers were as intelligible as native speakers to non-native listeners. This was the case whether the non-native speaker and listener shared the same language (matched interlanguage speech intelligibility benefit) or not (mismatched interlanguage speech intelligibility benefit). This interlanguage speech intelligibility benefit has not been consistently documented, suggesting that it is complex and influenced by various factors including the language combinations, as well as the proficiency of the non-native speakers (e.g., Stibbard & Lee, 2006) and the listeners (e.g., Imai, Walley, & Flege, 2005; Pinet, Iverson, & Huckvale, 2011).

If there is a non-native accent present in speech, this is easily perceived by listeners. The strength of the non-native accent impacts how others perceive the speaker. Native as well as non-native listeners judge speakers with non-native accents more negatively than those with native accents, e.g., as less fit for certain jobs (e.g., Kalin & Rayko, 1978; Roessel, Schoel, Zimmermann, & Stahlberg, 2019).

1.3 Lombard speech and non-natives

There are several studies that have researched non-native Lombard speech. Villegas and colleagues (2021) found that native Japanese speakers had higher sound pressure levels (a measure of loudness) in both their native Japanese and non-native English Lombard speech compared to plain speech, although the amount of increase depended on the combination of language and task. Investigating intensity with Chinese-English late-bilinguals (first language, L1, Chinese, second language, L2, English), Cai, Yin, and Zhang (2020) found that Lombard speech in the L1 and L2 had higher intensity compared to plain speech but the amount of increase depended on the language and noise condition combination. Investigating similar participants, Cai, Yin, and Zhang (2021) found that while the L1 and L2 speech showed an increase in intensity for Lombard speech, this increase was higher for the L2 than for the L1 speech in both the weak and strong noise conditions (30 and 60 dB SPL, respectively). Mok, Li, Luo, and Li (2018) examined mean intensity, mean F0, and duration of vowels in L1 Mandarin and L2 English plain and Lombard speech, finding that duration was longer and intensity was higher in L1 and L2 Lombard speech relative to plain speech. Further they found that mean F0 increased for two of the three tones studied in L1 Mandarin Lombard speech, while mean F0 decreased in L2 English Lombard speech compared to plain speech. Combined, these studies suggest that non-native speakers produce Lombard speech similarly to native speakers, while at times differing slightly depending on the task and acoustic measure. The research to date has been restricted to non-native English produced by speakers

of an Asian native language. Other native languages may exert other influences. The research into non-native Lombard speech should be expanded to other language combinations, to see if similar patterns emerge, investigating the same and more acoustic characteristics of Lombard speech.

There is limited research investigating the Lombard intelligibility benefit for non-native listeners. One study was conducted by Junqua (1993) and another by Cooke and García Lecumberri (2012). Junqua (1993) tested native speech with native (British and American-English) and non-native English (French) listeners, but did not find a Lombard intelligibility benefit for any of the listener groups. Cooke and García Lecumberri (2012) examined Spanish natives listening to native English plain and Lombard speech. Their findings showed that non-native listeners experienced a Lombard intelligibility benefit, but that it was slightly smaller than for the native listeners who were tested with the same materials (Lu & Cooke, 2008).

There seems to be a gap, as I am not aware of any published study having examined the perception of non-native Lombard speech. By examining the production and perception of non-native Lombard speech, we not only learn more about Lombard speech itself, but also about non-native speech production and perception in general.

1.4 Research Questions

This dissertation aims to identify and understand the characteristics of non-native Lombard speech, as this will lead to a better understanding of Lombard speech in general as well as of non-native speech production. For this, I chose to examine non-native Lombard speech using three distinct approaches; acoustics, intelligibility, and accentedness.

The research question concerning acoustics is: how does native and non-native Lombard speech potentially differ in acoustic measures and why? Non-native Lombard speech may differ from native Lombard speech because non-native speakers experience a higher cognitive load, which may influence their Lombard modifications. Additionally, native and non-native Lombard speech may differ because of the influence of the native language on the non-native Lombard speech. In order to answer this research question, I will examine potential differences between native and non-native English Lombard speech as well as the potential influence of the native language on non-native Lombard speech. This should give insight into the potential influence of the native language as well as the role of the cognitive demands of being non-native.

In terms of intelligibility, the research questions are: is there a Lombard intelligibility benefit for non-native speech? If so, how does the combination of the speakers' and listeners' native languages influence the size of the Lombard intelligibility benefit? Considering the extent to which the non-native speech elicits a Lombard benefit suggests the acoustic

modifications they may or may not be making.

Regarding accentedness, the research question is: how is non-native speech produced in different conditions evaluated in terms of accentedness compared to native speech? The accentedness evaluation of native and non-native speech informs us of the extent of the perceived change in pronunciation, and therefore also about the acoustic properties of non-native speech. Further, it may inform us about the potentially harmful consequences for non-native speakers when they produce speech with more effort in Lombard speech.

1.5 DELNN Corpus

At the center of this dissertation is the Dutch English Lombard Native Non-Native (DELNN) corpus, which I created in order to answer the research questions from this dissertation. It consists of native English speech produced by American-English speakers, non-native English speech produced by native Dutch speakers, and native Dutch speech produced by the same speakers. As far as I am aware, this is the first publicly available corpus that contains non-native Lombard speech. Comparing native and non-native Lombard and plain English speech allows us to determine whether there is a difference between native and non-native Lombard speech modifications, while comparing native Dutch and non-native English Lombard speech from the same speakers helps in understanding whether these potential differences are related to the influence of the native language or to the fact that the speaker is non-native. Dutch and English were chosen because of their similarities and differences as well as native Dutch speakers' relatively high proficiency in English.

Native Dutch and native American-English female speakers read English question-answer pairs, half in quiet, producing plain speech, and half in noise, producing Lombard speech. The native Dutch speakers further read question-answer pairs in Dutch. The question-answer pairs manipulated the location of contrastive focus in the answers. I intentionally included words with phonemes that were difficult for Dutch speakers of English, to examine potential pronunciation variation in their plain and Lombard speech.

All speakers were given the same material to read aloud, which results in the DELNN corpus having great consistency across speakers. This consistency allows for more comparable evaluations of acoustic measurements across speakers than if the speakers were asked to produce spontaneous speech. Moreover, this allows material from the DELNN corpus to be used for controlled experiments that compare Lombard with plain speech and native with non-native speakers. Tucker and Ernestus (2016) encourage researchers to utilize more ecologically valid casual speech. Nonetheless, I decided to use read speech in this early stage of research in non-native Lombard speech because of the reduced variability of read speech, which facilitates acoustic analyses and the selection of stimuli for highly controlled

perception experiments.

In addition to the speech recordings discussed above, the DELNN corpus includes phone and word level transcriptions of said speech. This facilitates the finding of specific phones and word tokens in the acoustic signal. The transcriptions will be beneficial to future researchers who choose to use the DELNN corpus.

1.6 Outline

This dissertation attempts to gain insight into the characteristics of non-native Lombard speech using three distinct perspectives; acoustics, intelligibility, and accentedness. The DELNN corpus, the basis of this entire dissertation, is discussed in detail in Chapter 2.

Chapters 2 through 6 are all related to the acoustics research question. **Chapter 2, 3, and 4** directly examine several acoustic measures: median F0, F0 range, spectral center of gravity, intensity, word duration, and voice onset time. Parts of Chapter 5 and Chapter 6 further address the acoustics research question by analyzing the speech signal using two computational measures. In **Chapter 5,** the High Energy Glimpsing Proportion metric is used and in **Chapter 6,** the forced aligner's transcription is used to investigate the production of phones that are difficult for the non-native speakers, to establish if their pronunciation of these phones changes from plain to Lombard speech.

Chapter 5 addresses the intelligibility research question, whether non-native speech elicits a Lombard intelligibility benefit and if so, how the size of the Lombard intelligibility benefit is affected by the combination of native and non-native English speech and native and non-native listeners. I tested two non-native listener groups – native Dutch and native Spanish – to see how the Dutch perform compared to the Spanish when listening to the non-native speech. This will be informative because the Dutch listeners share the native language of the non-native speakers, while the Spanish listeners do not, but Spanish does share some characteristics with English that Dutch does not.

Chapter 6 addresses the accentedness research question, by examining listeners' accentedness evaluations of native and non-native speech produced in two different conditions, not only differing by whether they represent plain versus Lombard speech but also in the presence versus absence of contrastive focus (contrastive focus on the target word produced in noise and no contrastive focus on the target word produced in quiet). If we assume that speakers put more effort (vocal and articulatory) into Lombard speech than plain speech, this might affect their perceived accentedness, which may in turn affect how they are evaluated by their listeners. I examine how the accentedness of non-native speech in these two conditions is evaluated compared to that of native speech in the same two conditions. The results of this study may have societal consequences.

In the final chapter, **Chapter 7,** I will discuss the combined findings from the three perspectives; acoustics, intelligibility, and accentedness. Together, these results will lead to conclusions about the characteristics of non-native Lombard speech. The chapter closely examines decisions made in the research process and discusses them as well as future research possibilities.

Chapter 2:

Acoustic characteristics of non-native Lombard speech in the DELNN corpus

Abstract

Lombard speech, speech produced in noise, has been extensively studied in native speakers, while non-native Lombard speech research has been limited to a few studies and several acoustic measures. The current article expands this research into non-native Lombard speech by investigating intensity, spectral center of gravity, word duration, and VOT. For this purpose, we compiled the Dutch English Lombard Native Non-Native (DELNN) corpus, which includes plain and Lombard speech from native English (American-English), non-native English (native Dutch), and native Dutch speakers. In the corpus, the location of contrastive focus is systematically varied, so that the difference between plain and Lombard speech can be compared between different prosodic constituents. The results of our analyses did not indicate differences in how native speakers of English and of Dutch adapt their English speech in noisy conditions, indicating that the non-native English are producing Lombard speech similarly to the native English. The comparison of the native Dutch and non-native English sentences produced by the same participants nevertheless suggests that, for all acoustic measurements except for word duration, the Dutch speakers adapt their speech differently in native Dutch than in non-native English. Combined, this would indicate that, when speaking English, Dutch speakers adapt their way of speaking in noisy conditions to native English.

This chapter is an edited version of:

Marcoux, K., & Ernestus, M. (submitted). Acoustic characteristics of non-native Lombard speech in the DELNN corpus.

2.1 Introduction

In noisy environments, such as train stations, canteens, and cafes, our way of speaking changes and is better understood in this environment, resulting in what is known as Lombard speech (Lombard, 1911). The specific acoustic modifications that occur when going from a quiet environment, where one produces plain speech, to a noisy environment, where one produces Lombard speech, have been thoroughly studied in native speakers. In contrast, there has been relatively little research dedicated to non-native speakers in noise. This study examines non-native Lombard speech acoustics and the potential influence of the native language.

Lombard speech is characterized by changes in acoustics compared to plain speech. These include but are not limited to an increase in fundamental frequency (F0), a widening of the F0 range, an increase in intensity, a shift in energy to higher frequency regions, and changes in duration (for a review see: e.g., Cooke, King, Garnier, & Aubanel, 2014). The extensive research on Lombard speech acoustics has been conducted in several languages including English (e.g., Lu & Cooke, 2008; Pisoni, Bernacki, Nusbaum, & Yuchtman, 1985; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988), Dutch (e.g., Bosker & Cooke, 2020), Spanish (e.g., Castellanos, Benedí, & Casacuberta, 1996), and French (e.g., Garnier & Henrich, 2014). All this research has focused on native speakers of these languages.

If Lombard speech is an automatic reaction to noisy environments, its properties may be language universal. For instance, increase in intensity may be assumed to depend on how much masking can be expected from the background noise on the speech, rather than depending on exactly which language is spoken. If Lombard speech adjustments are universal, one would expect no differences between native and non-native speakers of the same language. However, Zollinger and Brumm (2011) argue that while Lombard speech does elicit an involuntary response, it is "not a true reflex" (p. R614) since it can be modified by the speaker (e.g., producing a stronger effect when communicating than when in a non-communicative setting such as reading a text; e.g., Junqua, Fincke, & Field, 1999; Villegas, Perkins, & Wilson, 2021) and it can be inhibited with training (e.g., Pick, Siegel, Fox, Garber, & Kearney, 1989). Considering the speaker's ability to adapt their Lombard speech to the speaking conditions, we may observe more differences in Lombard speech, for instance, between languages.

If Lombard speech has language specific modifications, we would expect to observe differences when examining native and non-native Lombard speech. The native language influences many characteristics of plain non-native speech, including the quality of vowels, voice onset time (VOT) of consonants, and intonation, (e.g., Burgos, Cucchiarini, Van Hout, & Strik, 2013; Flege & Eefting, 1987; van Maastricht, Krahmer, & Swerts, 2016). Based on this, one may expect that non-native Lombard speech acoustics may also be influenced by the

native language.

In addition to the influence of the native language, non-native speech may be affected by the higher cognitive load non-native speakers experience relative to speaking a native language (Kormos, 2006; Segalowitz, 2010). Wester, García Lecumberri, and Cooke (2014) analyzed speech from native English and native Spanish speakers speaking in their native and non-native language (Spanish and English, respectively). They found that non-native speech was characterized by certain acoustic characteristics that are also present in hesitant speech, such as a slower speech rate and a smaller F0 range.

To our knowledge, non-native Lombard speech has only been investigated in a handful of studies in limited acoustic measures. Villegas, Perkins, and Wilson (2021) investigated native Japanese speakers producing native Japanese and non-native English speech. They found that both the native and the non-native speech showed an increase in sound pressure level in Lombard speech compared to plain speech, although the amount of increase depended on the combination of the native and non-native language with the task. Cai, Yin, and Zhang (2020) examined Chinese-English late bilinguals, finding that the speakers increased their intensity in noise in both Chinese (first language, L1) and in English (second language, L2), while the amount of increase depended on the language they spoke (L1/L2) in combination with the noise condition. The same researchers further investigated intensity with similar participants, finding that the L2 speakers increased their intensity more than L1 speakers in the two noise conditions (Cai, Yin, & Zhang, 2021). Mok, Li, Luo, and Li (2018) also examined speakers who had Mandarin as their first language and English as their second, examining mean intensity, F0, and duration of vowels. For the L1 Mandarin speech, they found longer durations, higher intensity and higher F0 (for two of the three tones studied) for vowels in noise. For the L2 English speech, they also found longer durations and higher intensity, but unexpectedly, lower F0 in noise than in quiet.

Chapters 3 and 4 will compare plain and Lombard speech as produced by native American-English speakers, by native speakers of Dutch speaking non-native English, and the same native speakers of Dutch speaking native Dutch. We will find that when going from plain to Lombard speech, the non-native speakers increase their median F0 and F0 range, in accordance with past research on native speech (F0; e.g., Pisoni, Bernacki, Nusbaum, & Yuchtman, 1985; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; F0 range: e.g., Garnier & Henrich, 2014; Welby, 2006). Chapter 3 will also observe an effect of the native language on the non-native language depending on the position of contrastive focus in the sentence. We shall observe that the Dutch non-native English speakers increase their median F0 when the sentence they read has contrastive focus early in the sentence, as they do in their native Dutch, while the native English speakers do not increase their F0 to a significant extent. Furthermore, for the late-focus sentences, these non-native English speakers have a

smaller F0 range increase compared to the native English speakers in their Lombard speech, but not as small as in their native Dutch (Chapter 4). We will interpret these results as an indication of a slight influence of the native language on the non-native Lombard speech.

In this article, we further explored the acoustic properties of non-native Lombard speech. Like the studies above, we investigated potential differences between native and non-native Lombard speech. Additionally, following Chapters 3 and 4, we examined how non-native Lombard speech relates to the speaker's native language.

Also like Chapters 3 and 4, we compared English sentences produced by native Dutch speakers with, on the one hand, the same sentences produced by native American-English speakers, and, on the other, with similar sentences produced by these same speakers in their native Dutch. We decided to focus on non-native English produced by Dutch native speakers since Lombard speech has been researched in both languages (English: e.g., Bosker & Cooke, 2018; Lu & Cooke, 2008; Pisoni et al., 1985; Van Summers et al., 1988; Dutch: e.g., Bosker & Cooke, 2020). Additionally, Dutch natives tend to have high proficiency in English, and are therefore likely to be able to adapt their speech to the environment, making them a good non-native population to study. Moreover, with both Dutch and English being Germanic languages, there are many similarities between them (e.g., post focus compression, non-tonal languages), making them easy to compare, while they still differ in many respects (e.g., voice onset time). Because Dutch Lombard speech has not been as extensively studied, this article also adds to our knowledge of native Dutch Lombard speech.

We first compared the difference between plain and Lombard speech in native and non-native English (the *English speech* comparison), to see whether the non-native speakers adapt their speech when going from plain to Lombard speech in the same way as the native English speakers do. We then compared the non-native English and native Dutch speech, produced by the same speakers (the *Dutch speakers* comparison). This comparison shows whether the Dutch adapt their speech, when going from plain to Lombard speech, in the same way in their native Dutch as in their non-native English. This second comparison may shed light on how to interpret the results from the English speech comparison. If no differences are found in the English speech comparison, the Dutch speaker comparison will show whether this is because the Dutch speakers have learnt how to produce Lombard speech in English (we then see a difference between native Dutch and non-native English) or because there are minimal to no differences between Dutch and English (we then see no difference between native Dutch and non-native English). If, in contrast, the English speech comparison shows differences between native English and non-native English, the Dutch speaker comparison may show whether these differences result from a transfer of Lombard properties from native Dutch to non-native English.

We examined four acoustic measures. The first characteristic of Lombard speech we investigated is intensity, which has been shown to increase compared to plain speech (e.g., Dreher & O'Neill, 1957; Junqua, 1993; Lu & Cooke, 2008; Pisoni et al., 1985; Van Summers et al., 1988). Considering that intensity and F0 are correlated (e.g., Gramming, Sundberg, Ternström, Leanderson, & Perkins, 1988), and that non-native Lombard English produced by Dutch native speakers may show F0 characteristics from Dutch (Chapter 3), we may expect differences in intensity between English Lombard speech produced by American-English and Dutch native speakers. These differences may be a function of the location of contrastive focus in the sentence, as they are for F0.

The second characteristic of Lombard speech we investigated is spectral Center of Gravity (CoG, as a measure of energy) of the utterance. As mentioned, Lombard speech is characterized by a shift in energy to higher frequency ranges (e.g., Pisoni et al., 1985; Van Summers et al., 1988). One way to measure the shift in energy is to examine spectral CoG, which has been shown to increase in English Lombard speech (e.g., Junqua, 1993; Lu & Cooke, 2008, 2009a). In plain speech, there are known differences between English and Dutch CoG for certain phones, such the /s/ in Dutch having a lower CoG than in English (see the references in Quené, Orr, & van Leeuwen, 2017). Because spectral CoG has not been investigated in native Dutch Lombard speech, we do not know whether there are differences between English and Dutch in this respect, and we cannot know whether non-native English produced by native Dutch speakers can show the signature of their native language with respect to spectral CoG. Nevertheless, we investigated this feature, as it seems to be an important characteristic of Lombard speech in general.

The third characteristic we examined is word duration. Previous research has indicated that words and sentences are mostly longer in Lombard than in plain speech (e.g., Dreher & O'Neill, 1957; Junqua, 1993; Van Summers et al., 1988). However, one study has found the opposite, documenting shorter sentence durations in Lombard speech, accompanied by shorter silence durations, which the researchers mention may be due to the speaker's urgency because of the noise (Varadarajan & Hansen, 2006). Because the differences in duration between Lombard and plain speech have not yet been investigated in Dutch to our knowledge, we do not know whether there are differences between English and Dutch in this respect, and, again, we cannot formulate predictions about whether non-native English produced by native Dutch speakers can show the signature of their native language with respect to duration.

The fourth acoustic measure we investigated is VOT (Voice Onset Time, the time from the release of the stop consonant as marked by a burst to the start of voicing; Lisker & Abramson, 1964). While the previous three acoustic measures have been extensively researched in Lombard speech, and have been shown to be different in plain and Lombard speech, this is not the case for VOT (no study investigates VOT in Lombard speech, to our knowledge). We

additionally examined VOT as Dutch and English plain speech differ in their VOT lengths. Lisker and Abramson (1964) reported average VOTs for English speakers as 58 ms for word initial /p/ and 80 ms for word-initial /k/ and for Dutch speakers, these values were 10 ms for /p/ and 25 ms for /k/. As a consequence, Lombard speech may affect VOT differently in Dutch than in English, which may surface in non-native English produced by native Dutch speakers. However, in examining native Dutch speakers producing English voiceless plosives, Simon and Leuschner (2010) found that trained and untrained post-secondary school Dutch students were producing longer VOTs in English than in Dutch and that these values were in native English speaker ranges. This suggests that the Dutch are able to adapt their voiceless VOTs to native English, at least in plain speech.

In order to investigate how Dutch speakers produce Lombard speech in non-native English, we built the Dutch English Lombard Native Non-Native (DELNN) corpus, which is the first speech corpus of non-native Lombard speech. The DELNN corpus consists of English plain and Lombard speech from native English speakers as well as from native Dutch speakers, who also produced native Dutch plain and Lombard speech. The 30 female native Dutch speakers and nine female native American-English speakers read contrastive questions-answer pairs, where the location of contrastive focus in the answers was manipulated. The corpus is available to other researchers and is described in detail in later sections.

We first describe the corpus in detail, followed by a section on the data we extracted from the corpus for the present acoustic study. In order to be able to investigate the effect of the presence versus absence (or location) of contrastive focus on word and phone based measures (word duration and VOT), we extracted data on these measures only from words that occur in both contrast conditions. As a consequence, all comparisons that we present of the acoustic characteristics could feature the type of speaker (either native American-English versus non-native English or non-native English versus native Dutch), speech style (plain and Lombard), as well as the presence / position of focus. We will thus be examining intensity and spectral CoG at the utterance level, duration at the word level, and VOT at the phone level to examine the role of nativeness in Lombard speech via the analysis of *English speech* and *Dutch speakers.*

2.2 DELNN Corpus

2.2.1 Speakers

The DELNN corpus includes the recordings from 39 female speakers, allowing us to have a homogenous sample. Males and females differ in some acoustic measures such as F0 and some research has reported slight differing pitch and energy changes in Lombard speech for women and men (e.g., Junqua, 1993). All speakers reported no vision or hearing issues nor

dyslexia or stuttering.

Of the 39 speakers in the DELNN corpus, 30 were native speakers of Dutch, with an average (*M*) age of 21.3 years, and nine were native speakers of American-English (*M* = 22.1 years). The Dutch natives were students at Radboud University, Nijmegen, The Netherlands (RU) and were completing their studies in Dutch. They all had native Dutch speaking parents and according to their LexTALE scores (*M* = 69.4, standard deviation (*sd*) = 15.8) (Lemhöfer & Broersma, 2012) on average they had a B2 English level proficiency as per the Common European Framework (Council of Europe, 2001). The American-English speakers were studying at RU at the time of recording and all had been residing in The Netherlands for less than a year and a half (ranging from 3 months to a year and a half). They were raised in the United States by at least one native English speaking parent.

2.2.2 Speech Materials

The speech materials consisted of question-answer pairs in which there was a target word embedded. For the English speech materials, there were three target word categories, chosen because of their difficulty for native Dutch individuals speaking English. The first category was formed by /θ/-initial words (e.g., *theater*). The /θ/ is problematic for native Dutch speakers as /θ/ does not exist in their phoneme inventory. They tend to produce other phones in its place, most often /t/ (Hanulikova & Weber, 2010). The second category of target words is English-Dutch cognates with a schwa in prestress position in American-English, which is represented by a <a> or <o> in the orthography, and that also corresponds to a full vowel in Dutch (e.g., *parade*). These schwa target words may pose difficulty as the Dutch may tend to produce the schwa as the full vowel that is present in the orthography and in their native Dutch. This may be especially so when the word receives contrastive focus, while the full vowel may be more likely to be correctly produced as schwa in non-accent position, due to vowel reduction in Dutch in this position (e.g., Booij, 1999). The final category consists of target words ending in voiced obstruents (e.g., *club*). These words are difficult for Dutch speakers since Dutch has final devoicing (e.g., Berendsen, 1986; Booij, 1985; Simon, 2010), meaning that voiced obstruents are produced as voiceless in syllable-final positions. Dutch speakers may also apply final devoicing to English words. Hereafter, these three categories are referred to as: /θ/, schwa, and voiced obstruent target words, respectively, as indicative of the problematic phoneme for the native Dutch speakers.

Each of the English target word categories consisted of 12 target words, resulting in a total of 36 English target words (see Appendix 1), with an average of 2.5 syllables per target word (*sd* = 1.0). For each target word, four question-answer pairs were created, resulting in a total of 144 English question-answer pairs. Each question had an average of 9.2 words

(*sd* = 1.1) and each answer an average of 9.3 (*sd* = 1.3). As an illustration, the four question-answer pairs for the target word *parade*, belonging to the schwa target word category, are presented below, where the speakers were instructed to emphasize the words in bold.

*1. Did the family go to the **beach** in Barcelona?*
*No, they went to the **parade** in Barcelona.*

*2. Did the **friends** go to the parade in Barcelona?*
*No, the **family** went to the parade in Barcelona.*

*3.Did Lily enjoy the flower **garden** in the spring?*
*No, she enjoyed the flower **parade** in the spring.*

*4.Did **Ellen** enjoy the flower parade in the spring?*
*No, **Lily** enjoyed the flower parade in the spring.*

As can be seen in the examples above, a target word appears in all answers of the question-answer pairs and in half of the questions. The location of contrastive focus was manipulated: half of the question-answers pairs had late-focus (examples 1 and 3), where the contrastive focus was on the target word, and the other half had early-focus (examples 2 and 4), where the target word was in a similar position as in the late-focus sentences, but the contrastive focus was earlier in the sentence, on a different word. The target words were never sentence final. By having the target words both in contrastive focus position and in non-focus position, the effect of contrastive focus position can be investigated.

The Dutch sentences were very similar in structure to the English sentences except that instead of having 36 target words for a total of 144 question-answer pairs, the Dutch had 24 target words resulting in 96 question-answer pairs. Twelve of these target words were the Dutch translations of the English schwa words and for the other 12 target words, nouns were chosen (see Appendix 2). The Dutch target words were on average 3.0 syllables in length (*sd* = 0.7). Each question had an average of 7.5 words (*sd* = 0.7), and each answer an average of 8.2 (*sd* = 0.8). Below are the four Dutch question-answer pairs for the Dutch target word *parade*. Early-focus can be found in 5 and 7, and late-focus in 6 and 8, respectively. As mentioned, the late-focus condition indicates that the contrastive focus occurs later in the sentence, on the target word.

*5. Zagen de **jongens** de parade gisteren?*
*Nee, de **studenten** zagen de parade gisteren.*
*'Did the **guys** see the parade yesterday?*
*No, the **students** saw the parade yesterday.'*

*6. Zagen de studenten de **voorstelling** gisteren?*
*Nee, ze zagen de **parade** gisteren.*
*'Did the students see the **performance** yesterday?*
*No, they saw the **parade** yesterday.'*

*7. Bezochten de **buren** de parade vanmiddag?*
*Nee, de **kinderen** bezochten de parade vanmiddag.*
*'Did the **neighbors** visit the parade this afternoon?*
*No, the **children** visited the parade this afternoon.'*

*8. Bezochten de kinderen de **speeltuin** vanmiddag?*
*Nee, ze bezochten de **parade** vanmiddag.*
*'Did the children visit the **playground** this afternoon?*
*No, they visited the **parade** this afternoon.'*

2.2.3 Lists

From these question-answer pairs, three main lists were created for each language. Every list contained all question-answer pairs of the given language. The first half of each list was produced as plain speech and the second half as Lombard speech. This led to four blocks, early-focus plain, late-focus plain, early-focus Lombard, and late-focus Lombard. As described above, four question-answer pairs were created per target word, one for each of these blocks. The four question-answer pairs consisted of two matched question-answer pairs, of which one of the pairs in the matched question-answer pairs is early- and one is late-focus, for example (1)-(2) and (3)-(4). In the lists, the matching pair of question-answer pairs both occurred in either the plain or the Lombard blocks. Apart from this matching criterion, the order of question-answer pairs was random.

For each language, for counterbalance purposes, an additional three lists were created from the three main lists. These additional lists had the question-answer pairs that were in Lombard speech as plain speech, and vice versa. Furthermore, the order of the early- and late-focus blocks were flipped within the plain and Lombard conditions. As a result, these additional lists adhered to the same criteria as the main lists: plain precedes Lombard and the matching question-answer pairs occurred in the same half. A filler was added as the first item of each block.

This resulted in six lists of 144 English question-answer pairs and four fillers and six lists of 96 Dutch question-answer pairs with four fillers. This formation of lists means that each speaker produced different question-answer pairs in the plain and Lombard conditions.

2.2.4 Procedure

All recordings were made at Radboud University, in a sound attenuated room. During the recording session, participants wore Sennheiser HD 215 MKII DJ headphones. Nothing was played via the headphones during the quiet condition, while speech shaped noise (SSN) was played at 83 dB SPL (77 dBA) during the noise condition, to elicit Lombard speech. The SSN level output was calibrated using the Brüel & Kjær Type 4153 artificial ear. In the recording session, participants were asked to place emphasis on the words in bold and to read at their own pace. They had a break after each block.

In the recording session participants sat 15 cm from the microphone, which was connected to an AudiTon amplifier, and in turn to a steady state recorder, where the recording was saved. The AudiTon amplifier was used to get the highest quality recordings without peaking and was located outside the recording booth so that the researcher could adjust it, without disturbing the participant. When this amplifier was adjusted, a beep, not audible to the participant was elicited, which can be used to calibrate with the microphone's sensitivity and calculate the normalized intensity of the recording. We used two different models of microphones, the Sennheiser ME 64 and the Sennheiser ME 65. The Sennheiser ME 64 has a sensitivity of 31 mV/Pa, corresponding to 70.2 dB, while the Sennheiser ME 65's sensitivity is 10 mV/Pa, corresponding to 80 dB.

Since the intensity of the participants' voices was expected to vary between plain and Lombard speech, we adjusted the AudiTon before the plain speech block and the Lombard speech block. Therefore, before the first (plain) and third (Lombard) blocks, participants read a short passage so that the AudiTon amplifier could be calibrated to the highest loudness without peaking. Furthermore, each block started with a filler question-answer pair, in case further calibration was needed. If the calibration was not ideal and the speaker was too close to peaking, we calibrated as needed during the session, unbeknownst to the participants.

After recording the English stimuli, the Dutch participants completed the LexTALE task (Lemhöfer & Broersma, 2012), in which they were presented with 60 English words and non-words on the screen and had to indicate for each whether it was a real English word. This task indicates the individual's overall English proficiency level as per the European Common Framework (Council of Europe, 2001). The Dutch participants returned within a week to record the Dutch stimuli, which followed the same recording procedure. We decided to always present the English stimuli in the first session and the Dutch in the second session, as we did not want participants to drop out when they learned that the following session

would be in a foreign language. This meant that the session of the language was confounded with the order of the session.

All participants completed a language background questionnaire. Additionally, all participants gave informed consent and were compensated upon completion of the recording session(s) with course credit or gift vouchers. In total, the English recording session lasted approximately an hour and the Dutch session, about 45 minutes.

2.2.5 Word and phone level transcriptions

The speech recordings were aligned at the word and phone level using the Montreal Forced Aligner (MFA: McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017) for the English data and Kaldi (Povey, Boulianne, Burget, Motlicek, & Schwarz, 2011) for the Dutch data. Details on the alignment process can be found in Appendix 3. These word and phone level transcriptions were used for the acoustic measures; the silences were removed from the sentences before measuring intensity and spectral CoG, word durations were measured on the word level transcriptions, and the VOT measures were annotated by humans who used the transcriptions to orient themselves.

The DELNN corpus is available to researchers upon request. This includes the speech recordings as well as the Praat TextGrids with the word and phone level transcriptions.

2.3 Methods for the acoustic measures

2.3.1 Materials

The materials for the acoustic measures only consisted of the answers from the question-answer pairs, not including the filler trials. Because the non-native participants especially had difficulties with question-answer pairs that contained the target words *massage*, *thermodynamics*, and *thermometer*, we eliminated answers with these target words for all four acoustic measures. This meant that all acoustic measures were calculated for maximally 132 English stimuli per participant (144 total sentences minus the three difficult target words produced in the four blocks) and an additional 96 Dutch stimuli for the native Dutch speakers. Further, in order to ensure that there was no clipping in the sound files, utterances with maximums or minimums greater than 1.00 Pa or less than -1.00 Pa, as indicated by Praat (version 6.0.37; Boersma & Weenink, 2018), respectively, were excluded from the intensity analysis. This resulted in the exclusion of another 125 answers for the analyses of intensity.

We analyzed word duration and VOT measures of the target words, because only these words informed us about the role of plain and Lombard speech in combination with the effect of contrastive focus. Because speakers produced another real word instead of the target word

in three instances and because of a technical error in the alignment process, we lost a total of four target word tokens for the analysis of word duration.

For the analysis of VOT, we only focused on word-initial /p/ or /k/ target words followed by either a vowel or /r/ or /l/ in English, since these phones allowed for a clear onset of voicing. For Dutch stimuli, the /p/ and /k/ had to be followed by a vowel, and not by an /r/. This was due to the fact that there is great variation in Dutch /r/ pronunciations, which does not provide a consistently clear onset of voicing and is therefore problematic for VOT measurements. This left us with *cab*, *cadaver*, *computer*, *club*, *crib*, *parade*, *police*, *pub*, and *professor* in English. For the Dutch stimuli, we extracted VOT values from *computer*, *kadaver*, *kostuum*, *parade*, and *politie* (in English: *computer*, *cadaver*, *costume*, *parade*, and *police*). We excluded 12 tokens where the voiceless plosive was followed by a voiceless vowel, the vowel was absent, or where there was prevoicing or frication as we were unable to accurately measure the VOT in these tokens.

2.3.2 Procedure

To calculate mean intensity values (in dB) of the answers as well as of the calibration beeps, we used Praat's (version 6.0.37; Boersma & Weenink, 2018) command "To Intensity" with a pitch floor of 100 and an auto time step. In calculating intensity, the "subtract mean" option was not implemented, as offset was minimal. The dB averaging method was used. In order to calculate the normalized intensity of each utterance, the difference between the microphone's sensitivity and the intensity of the closest calibration beep was added to the intensity of the utterance. For example, if the Sennheiser ME 65 was used to record an utterance, and the utterance was 65dB (80dB sensitivity of the microphone - 70 dB of the beep = 10dB) 10 dB was added to the utterance intensity to get the normalized intensity of the utterance (75 dB). We obtained spectral CoG values over the entire utterance by converting the sound file to spectrum using the slow setting in Praat and then computing the center of gravity in the power spectrum, in Hertz (Hz).

The durations (word durations and VOT) were calculated in Praat (version 6.0.37; Boersma & Weenink, 2018) by subtracting the start time from the end time of each token and the values were converted to milliseconds. If the target word was produced multiple times in one utterance, we took the first instance, even if this occurred just before a restart. For the VOT measurements, three annotators, trained by the authors, marked the VOT from the start of the burst to the start of the periodicity (at the zero-crossing boundaries), as is illustrated in Appendix 4. In order to see whether VOT duration only varies because of an overall lengthening of the speech materials, corresponding to slower speech rate, we included a durational measure as a control predictor in the VOT analyses. The annotators therefore

also marked the end of the vowel (which included an intervening liquid in the case of *club*, *crib* and *professor*) so we could calculate the duration from the end of the VOT to the end of the vowel. The end of the vowel was chosen rather than the whole target word as vowels and consonants are lengthened differently in Lombard speech, with vowels being elongated more than consonants (e.g., Castellanos et al., 1996; Garnier & Henrich, 2014; Junqua, 1993). Inter-rater agreement among the annotators is described in Appendix 5.

2.3.3 Analysis

For each acoustic measure, two separate analyses were conducted. One investigated *English speech*, comparing the native and the non-native English data. The other examined *Dutch speakers*, analyzing the same speakers, producing native Dutch and non-native English data. The two analyses are presented separately in the results sections.

Our predictors of interest for the intensity, spectral CoG, duration, and VOT analyses for the *English speech* and the *Dutch speakers* analyses were Speech Style (plain, Lombard), Speaker Nativeness (native, non-native) and Focus (at the sentence level, focus indicates early- or late-focus, at the word level focus indicates whether the word received contrastive focus or not). The crossed-random intercepts were Speaker and Answer for intensity and spectral CoG, and Speaker and Target Word for duration and VOT.

We included scaled and centered Trial Number as well as scaled and centered Occurrence (block number) as control variables. In some cases, there were technical issues and the experimenter asked the participant to redo the affected stimuli, resulting in Trial Numbers higher than 144 and Occurrences higher than four.

The VOT analysis had the following additional control variables: Plosive (p, k), Previous Phone (voiced, voiceless, or silence), Syllable Stress, and Duration (from the end of the VOT to the end of the vowel). The control predictor Plosive was included since /p/ and /k/ have been shown to have different VOT lengths (e.g., Lisker & Abramson, 1964). Since context may influence VOT (e.g., Yao, 2009), we also considered Previous Phone as a control predictor. Syllable Stress, indicating whether the syllable was stressed, was included as it has been reported to also affect VOT length (e.g., Cho & McQueen, 2005; Lisker & Abramson, 1967; Simonet, Casillas, & Díaz, 2014). The control predictor Duration, was included based on the results from Hazan, Grynpas, and Baker (2012), who found that increased word duration, corresponding to slower speech rate, explained the longer VOT for /p/ in clear speech.

We used R (version 3.5.1; R Core Team, 2016) to perform linear mixed effects models from the *lme4* package (version 1.1.21; Bates, Mächler, Bolker, & Walker, 2015), using the Nelder-Mead optimizer as it led to the best convergence. Before beginning the analysis, we first removed outliers, defined as 2.5 standard deviation above or below the grand mean.

We then began with a hypothesis-based model, which included interactions among the predictors of interest (Speech Style, Nativeness and Focus) and simple effects for the control variables. If a predictor was not significant and not in a significant interaction ($t < 1.96$), then it was removed from the model. Once everything in the fixed structure was set, we proceeded to the random structure. We then checked whether the addition of random slopes improved the models' AICs, using anova(). If an addition did not lead to an improvement, it was not included. If the fitting produced a warning, we did not proceed with that model. Once the random structure was finalized, we removed the data points that resulted in absolute standardized residuals above 2.5 and refitted the model. We checked this refitted model to ensure that all fixed predictors included were significant, using the summary() function as well as the Anova() function from the *car* package (version 3.0.6; Fox & Weisberg, 2019). Additionally, the residuals of each model were examined via a histogram as well as a qqplot to ensure that they were normally distributed, which was the case with all models reported below. In the instance that there was a three-way interaction in the refitted model among the predictors of interest (Speech Style * Nativeness * Focus), as was the case with the intensity data, we split by Focus to better understand the data. In the case of the split model, we started from the model with the three-way interaction and removed non-significant interactions and simple effects until only significant interactions and simple effects remained (without re-entering simple effects or interactions that were not significant in the model with the three-way interaction).

Plain speech (Speech Style), early-focus or not contrastive focus (Focus), and non-native speaker (Speaker Nativeness) were on the intercept. We chose plain speech and early-focus as we consider them our baseline. In order to compare the effects of predictors between the *English speec*h data and the *Dutch speakers*' data, we had non-native speakers on the intercept as the non-native speakers appear in both comparisons. The plots were created using the *ggplot2* (version 3.2.1; Wickham, 2016) package.

2.4 Results

2.4.1 Intensity

The intensity data are shown in Figure 1. The corresponding models comparing native and non-native English (*English speech*) and comparing non-native English and native Dutch (*Dutch speakers*) intensity values are presented in Table 1.

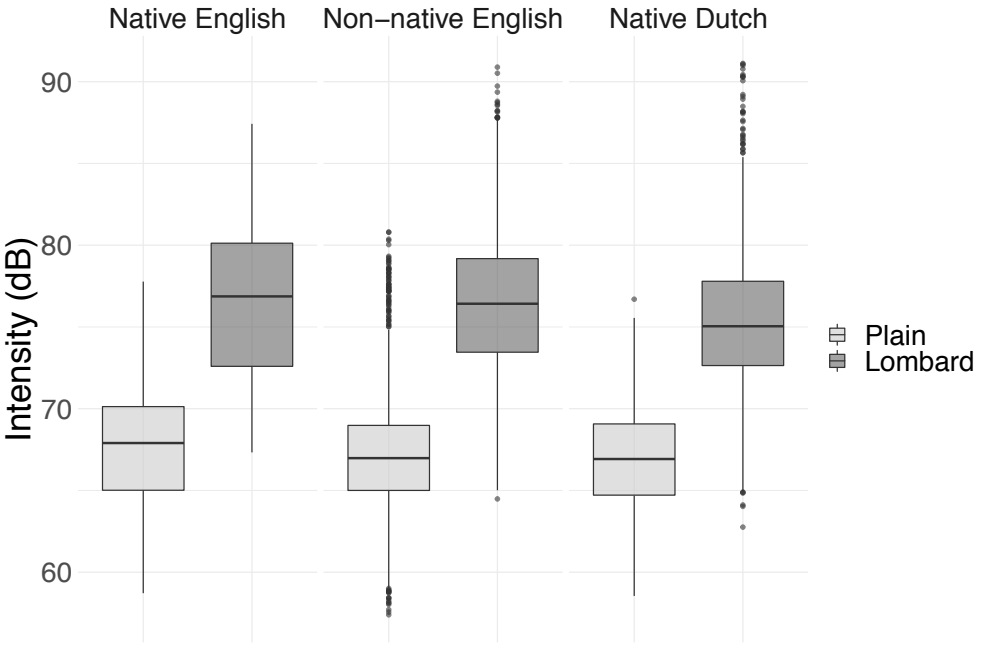

Figure 1: *Average intensity data for native English, non-native English, and native Dutch split by speech style. A box indicates the upper quartile, median, and lower quartile from top to bottom respectively. The ends of the vertical lines indicate the minimum and maximum, respectively, excluding the potential outliers, which are indicated by the dots. Figure 1 visualizes the data before the outliers were removed for the statistical analysis.*

Table 1: *lmer models of native and non-native English (English speech) intensity and non-native English and native Dutch (Dutch speakers) intensity split by Focus position.*

	English speech		Dutch speakers			
			Early-Focus		Late-Focus	
Fixed Effects	*β*	*t*	*β*	*t*	*β*	*t*
Intercept	67.4	128.7	67.6	128.7	67.5	111.2
Speech Style: Noise	8.3	14.9	7.2	11.3	8.2	13.1
Focus: Late	0.3	1.4	NA	NA	NA	NA
Nativeness: Native	-	-	-	-	-0.2	-0.4
Trial Number	0.3	9.1	0.8	11.6	0.4	4.4
Speech Style: Noise * Focus: Late	0.4	4.2	NA	NA	NA	NA
Speech Style: Noise * Nativeness: Native	-	-	-	-	-0.3	-3.3
Random Effects		*SD*		*SD*		*SD*
Answer (Intercept)		1.0		1.0		1.2
Speech Style by Answer		0.4		-		-
Speaker (Intercept)		3.1		3.2		3.2
Speech Style by Speaker		3.3		3.4		3.2
Nativeness by Speaker		-		1.9		2.1
Residual		1.1		1.0		1.1

2.4.1.1 English speech

For the fitting procedure for the model comparing the native with the non-native English intensity data, we removed 17 data points, which resulted in 4901data points being analyzed. The model (see first two columns in Table 1) revealed a significant simple effect of Speech Style which was modulated by Focus. Together these effects indicate that intensity increased significantly for Lombard speech compared to plain speech and that the increase was even larger for sentences with late-focus. Additionally there was a significant effect of Trial Number, with intensity increasing as the recording session progressed. The random structure showed that the intensity differed per Answer as well as per Speaker (as indicated by the significant random intercepts of Answer and Speaker). Moreover, the effect of Speech Style differed by Answer and by Speaker (as shown by the significant random slopes of Speech Style per Answer and Speech Style per Speaker).

2.4.1.2 Dutch speakers

The fitting procedure for the model comparing the non-native English and native Dutch speech produced by the same participants implied the removal of 61 data points resulting in a total of 6411 data points being analyzed. There was a three-way interaction in the model (Speech Style * Nativeness * Focus: β = -0.3, t = -3.0) and we therefore split the data by Focus to better understand the effects of Speech Style and Nativeness (see Table 1). For the early-focus sentences, the model revealed significant simple effects of Speech Style and Trial Number (see third and fourth columns of Table 1). Lombard speech had higher intensity than plain speech and, as the recording session progressed, intensity increased as well. For the late-focus sentences, there was a significant effect of Speech Style that was modulated by Nativeness. Lombard speech had higher intensity than plain speech but slightly less so in native Dutch. Trial number was also significant for the late-focus sentences; intensity increased over the recording session.

The random structures for both the early- and late-focus sentences revealed that intensity differed per Answer and per Speaker (as shown by the significant random intercepts of Answer and Speaker). Additionally both the effect of Speech Style and the effect of Nativeness varied per Speaker (as shown by the significant random slopes of Speech Style per Speaker and Nativeness per Speaker).

2.4.1.3 Intensity Interim Discussion

Lombard speech had higher intensity than plain speech in all speech, native English, native Dutch, and non-native English. Further, as the experiment progressed, intensity increased for all speakers.

For the *English speech* (native and non-native English data), in addition to the effect of Lombard speech having higher intensity, we observed that the Lombard sentences with late-focus were produced with even higher average intensity than the Lombard early-focus sentences. In English and Dutch, material after the focus undergoes post-focus compression (PFC, English: e.g., Cooper, Eady, & Mueller, 1985; Xu & Xu, 2005, Dutch e.g., Hanssen, Peters, & Gussenhoven, 2008; Rump & Collier, 1996), with a lowering and narrowing of the F0 range (e.g., Xu, 2011) as well as a lowering of intensity (e.g., Chen, 2015). Considering that intensity decreases after the focus, it is of no surprise that the Lombard late-focus sentences had a larger increase in intensity than the Lombard early focus sentences, as less material underwent PFC, allowing for a larger increase.

When examining *Dutch speakers* producing native Dutch and non-native English, we found a different pattern. For the late-focus sentences, the speakers produced lower intensity in their native Dutch than in their non-native English in Lombard speech. This is not likely to be due to stimuli differences in the two languages because we do not observe a general effect

of nativeness, both in plain and Lombard speech, but rather only in Lombard speech.

Hence, while all speakers increased their intensity when speaking in noise, when speaking Dutch, the Dutch natives did less so for late focus-sentences. As the Dutch speakers did not do the same in non-native English, apparently, the Dutch speakers appropriately adapted their way of producing Lombard speech when speaking non-native English.

2.4.2 Spectral CoG

The spectral CoG data for all speakers – native English, non-native English, and native Dutch – are visualized in Figure 2. Table 2 presents the corresponding models comparing native and non-native English (*English speech*) and comparing non-native English and native Dutch (*Dutch speakers*) in spectral CoG.

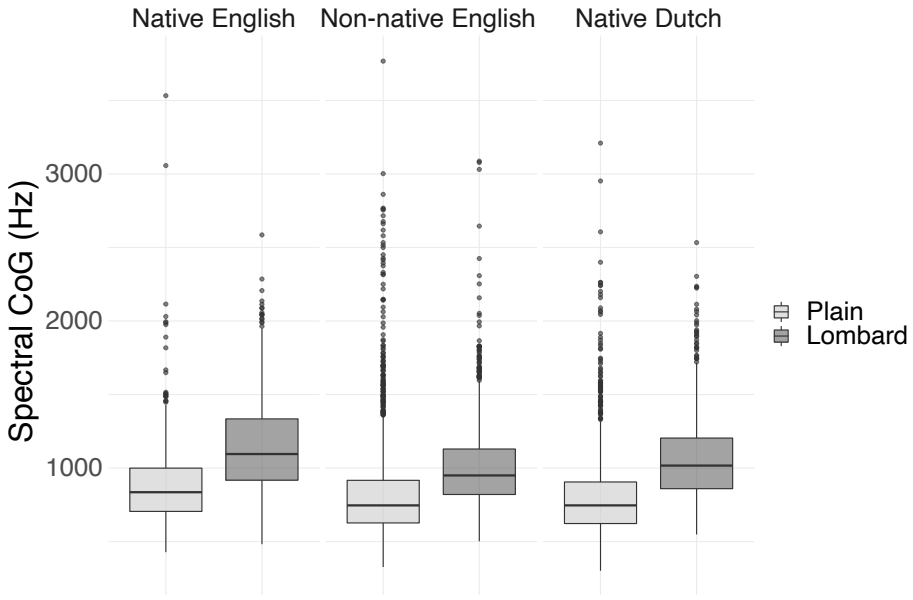

Figure 2: *Spectral CoG data for native English, non-native English, and native Dutch split by speech style. A box indicates the upper quartile, median, and lower quartile from top to bottom respectively. The ends of the vertical lines indicate the minimum and maximum, respectively, excluding the potential outliers, which are indicated by the dots. Figure 2 visualizes the data before the outliers were removed for the statistical analysis.*

Table 2: *lmer models of native and non-native English (English speech) spectral CoG and non-native English and native Dutch (Dutch speakers) spectral CoG.*

	English speech		Dutch speakers	
Fixed Effects	*β*	*t*	*β*	*t*
Intercept	807.9	30.8	785.7	30.3
Speech Style: Noise	216.8	8.8	200.8	9.1
Nativeness: Native	-	-	-2.0	-0.1
Speech Style: Noise * Nativeness: Native	-	-	50.9	5.6
Random Effects		*SD*		*SD*
Answer (Intercept)		132.7		141.8
Speech Style by Answer		62.2		51.0
Speaker (Intercept)		146.5		124.2
Speech Style by Speaker		149.4		116.2
Residual		111.9		120.8

2.4.2.1 English speech

The fitting procedure for the model comparing the native with the non-native English data implied the removal of 127 outliers for the analysis of 5020 data points. The first two columns in Table 2 above list the results of the model. The model revealed that the only significant predictor of interest was Speech Style, indicating that spectral CoG values increased for Lombard speech compared to plain speech for both native and non-native English to a similar extent. The random structure revealed that the spectral CoG varied per Answer and per Speaker (shown by the significant random intercepts of Answer and of Speaker). Additionally the effect of Speech Style differed per Answer and per Speaker (significant random slope of Speech Style by Answer and of Speech Style by Speaker).

2.4.2.2 Dutch speakers

For the fitting procedure for the model comparing the non-native English and native Dutch speech produced by the same participants, we removed 140 outliers, which resulted in a total of 6700 data points being analyzed. The last two columns of Table 2 above list the results of the model. The statistical model revealed simple effects of Speech Style as well as an interaction of Speech Style with Nativeness. Together these effects indicate that spectral CoG increased for Lombard speech, and this increase was larger in Dutch than in English. The random structure is similar to the *English speech* spectral CoG model, with spectral CoG

varying per Answer and per Speaker as well as differing for Speech Style per Answer and per Speaker.

2.4.2.3 Spectral CoG Interim Discussion

Our analyses showed that spectral CoG was higher in Lombard speech than in plain speech. The size of the Lombard effect seems similar for the native and non-native speakers of English but to be larger for native Dutch. This suggests that in regards to spectral CoG, the Dutch speakers were doing something slightly different in their Lombard speech in their native Dutch than in their non-native English. This suggests that the Dutch natives were adapting their spectral CoG to the native English speakers when producing non-native English Lombard speech. Note that, as for intensity, it is unlikely that the difference between Lombard speech in native Dutch and in non-native English can be ascribed to the differences in stimuli. If that were the case, we would have found a difference for plain speech as well.

2.4.3 Duration

The data for the durations of the target words in milliseconds are visualized in Figure 3 below. The corresponding models comparing native and non-native English (*English speech*) and comparing non-native English and native Dutch (*Dutch speakers*) target word duration are shown in Table 3. For the former model, we removed 63 outliers leaving 5081 data points, and, for the latter model, we removed 87 outliers leaving 6749 data points to be analyzed.

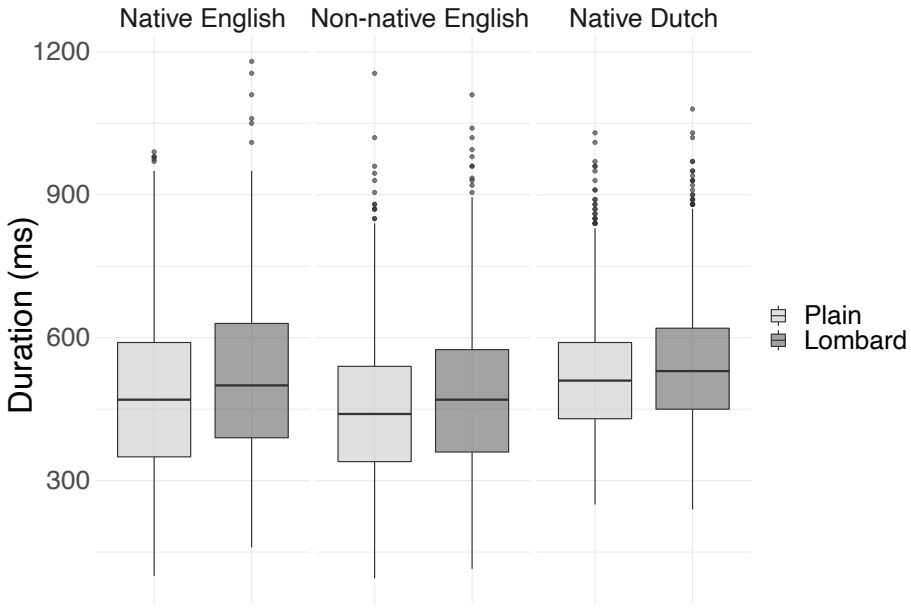

Figure 3: *The durations of target words produced by native English, non-native English, and native Dutch split by Speech Style. A box indicates the upper quartile, median, and lower quartile from top to bottom, respectively. The ends of the vertical lines indicate the minimum and maximum, respectively, excluding the potential outliers, which are indicated by the dots. Figure 3 visualizes the data before the outliers were removed for the statistical analysis.*

Table 3: *lmer models of native and non-native English (English speech) and non-native English and native Dutch (Dutch speakers) target word durations.*

	English speech		Dutch speakers	
Fixed Effects	*β*	*t*	*β*	*t*
(Intercept)	382.6	17.3	399.5	20.5
Speech Style: Noise	68.5	15.0	31.9	5.3
Nativeness: Native	5.3	0.2	71.5	2.7
Focus: Contrastive	81.8	43.6	78.7	14.0
Occurrence	-25.2	-14.6	-8.1	-2.7
Nativeness: Native * Focus: Contrastive	64.6	16.4	-	-
Speech Style: Noise * Focus: Contrastive	-	-	8.6	3.4
Random Effects		*SD*		*SD*
Speaker (Intercept)		61.0		46.3
Speech Style by Speaker		19.1		13.3
Focus by Speaker		-		29.3
Target word (Intercept)		109.4		99.6
Nativeness by Target word		27.7		-
Residual		58.1		51.1

2.4.3.1 English speech

The statistical model for native and non-native English revealed a significant simple effect of Speech Style, indicating that the duration of the word increases in Lombard speech. Additionally, we observe a significant effect of Focus (contrastive focus) and an interaction between Focus and Nativeness. Together this suggests that when a word receives contrastive focus, the word duration is longer, and that when native English speakers produce it, it is differentially longer. Finally, we observed a significant effect of Occurrence, indicating that the subsequent occurrences of a target word were shorter. The random structure revealed that duration differed per Speaker and per Target Word (as shown by the significant random intercepts of Speaker and Target Word) and that the effect of Speech Style differed by Speaker while the effect of Nativeness differed by Target Word (as shown by the significant random slopes of Speech Style by Speaker and Nativeness by Target Word).

2.4.3.2 Dutch speakers

From the model comparing non-native English and native Dutch we observed significant simple effects of Speech Style and Focus, which also interact with each other. The target words were longer in Lombard speech compared to plain speech and longer if carrying contrastive focus. If the target word was produced in Lombard speech with contrastive focus, the duration was even longer. Additionally we observed a significant simple effects of Nativeness and Occurrence, with the non-native English speakers producing shorter target words, and subsequent occurrences of the target word being shorter. The random structure indicates that word duration varied per Target Word and per Speaker (significant random intercepts of Speaker and Target Word) and that the effects of both Speech Style and Focus varied per Speaker (significant random slopes of Speech Style by Speaker and Focus by Speaker).

2.4.3.3 Duration Interim Discussion

All sets of duration data showed several similar patterns. Most importantly for our research questions, the target words produced as Lombard speech were longer than those produced as plain speech. Additionally, subsequent productions of target words were shorter in duration, in line with past research (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Fowler & Housum, 1987). Also in line with previous research, target words with contrastive focus had increased durations (e.g., Cooper et al., 1985).

There were also differences among the word duration datasets. In examining the target word durations in *English speech*, the native English speakers showed a larger effect of focus than the non-native English speakers. The *Dutch speakers* (native Dutch and non-native English speech) showed a larger effect of focus in Lombard than in plain speech, but due to the absence of a three way interaction of Nativeness with Speech Style and Focus for the *English speech* dataset, it is unclear whether the Dutch speakers differ in this respect from the native English speakers. Finally, the dataset of *Dutch speakers* show an effect of nativeness, with Dutch target words having longer durations. This is most likely due to differences between the Dutch and English stimuli, with the Dutch stimuli having more syllables on average per target word (Dutch: $M = 3.0$, $sd = 0.7$, English: $M = 2.2$, $sd = 1.0$).

2.4.4 VOT

The VOT data for native English, non-native English, and native Dutch are presented in Figure 4. The statistical models for native and non-native English (*English speech*) and for the non-native English and native Dutch (*Dutch speakers*) are shown in Table 4 below. For

the models, we removed 41 and 50 outliers leaving 1358 and 1618 data points to be analyzed, respectively.

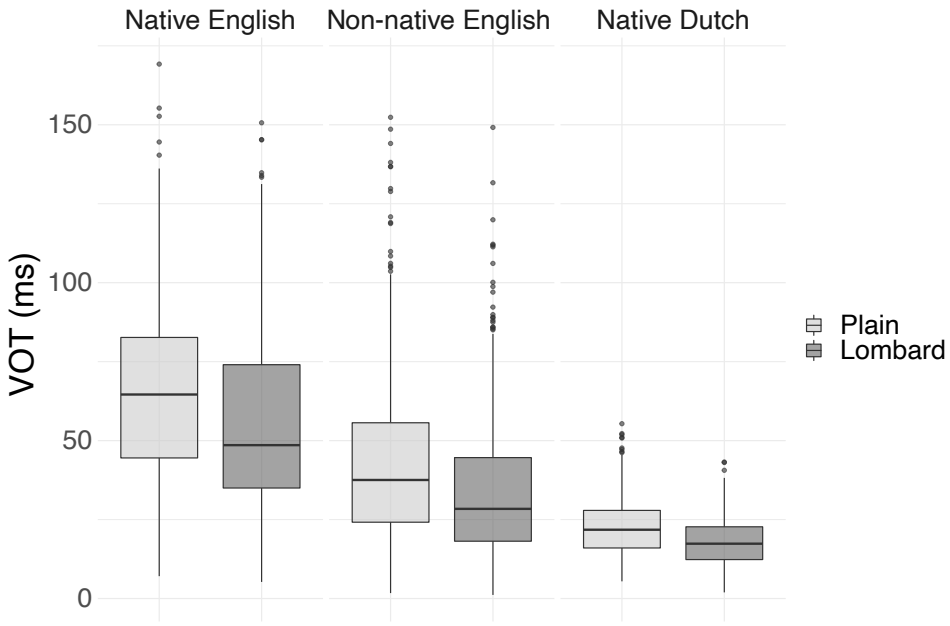

Figure 4: *The VOT of /p/ and /k/ produced by native English, non-native English, and native Dutch split by Speech Style. A box indicates the upper quartile, median, and lower quartile from top to bottom, respectively. The vertical ends of the lines indicate the minimum and maximum, excluding the potential outliers, which are indicated by the dots. Figure 4 visualizes the data before the outliers were removed for the statistical analysis.*

Table 4: *lmer models of native and non-native English (English speech) and non-native English and native Dutch (Dutch speakers) VOT.*

	English speech		Dutch speakers	
Fixed Effects	*β*	*T*	*β*	*t*
Intercept	35.8	6.7	46.0	9.4
Speech Style: Noise	-5.8	-2.7	-7.4	-8.5
Nativeness: Native	12.4	3.7	-15.8	-2.6
Focus: Contrastive	3.2	3.7	2.9	4.6
Plosive: p	-	-	-12.6	-2.2
Previous phone: Silence	-	-	-4.6	-2.4
Previous phone: Voiced	-	-	-2.0	-1.1
Duration	39.5	3.9	-	-
Occurrence	-2.1	-2.8	-	-
Speech Style: Noise * Nativeness: Native	-	-	3.2	3.1
Nativeness: Native* Focus: Contrastive	12.4	6.8	-4.4	-4.3
Random Effects		*SD*		*SD*
Speaker (Intercept)		9.7		5.6
Speech Style by Speaker		5.1		3.2
Target Word (Intercept)		14.9		10.8
Speech Style by Target Word		3.7		-
Residual		13.5		10.0

2.4.4.1 English speech

In analyzing native and non-native English VOT data, we found significant simple effects of Speech Style, Occurrence, and Duration. The VOTs were shorter when produced as Lombard speech compared to plain speech, following occurrences of the target word had shorter VOTs, and as the following segment duration increased, the VOT was longer. Additionally, there were simple effects of Nativeness and Focus, which interacted with each other. The VOT was longer if the speaker was a native English speaker and if the target word received contrastive focus, and it was lengthened further if both were the case. The random structure showed that VOT differed per Speaker and per Target Word (significant intercepts of Speaker and Target

Word) and that the effect of Speech Style varied by Speaker and that of Speech Style by Target Word (significant slopes of Speech Style by Speaker and Speech Style by Target Word).

2.4.4.2 Dutch speakers

From the model on non-native English and native Dutch speech, we see significant simple effects of Speech Style, and Nativeness, as well as an interaction of the two. The simple effects show that, in plain speech (the condition at the intercept), the native Dutch speakers shortened their VOT in Dutch compared to non-native English and that in English (again the condition at the intercept) they shortened their VOT in Lombard speech compared to non-native plain speech. The interaction of Speech Style and Nativeness shows that the effect of Speech Style on VOT is smaller in Dutch. In order to observe whether Speech Style effected VOT for native Dutch speech, we releveled the model with native Dutch speakers on the intercept. This releveled model indicated that this is the case (β_{Noise} = -4.2, t = -4.2). The Dutch thus also shortened their VOT when going from plain to Lombard speech in their native Dutch, although to a lesser extent than in non-native English, as indicated by the interaction in Table 4.

We also found an effect of Focus and an interaction of Nativeness with Focus. Table 4, with non-native English on the intercept, shows that contrastive focus in non-native English lengthened VOT. A releveled model, with native Dutch on the intercept, was used to determine whether the effect of Focus was significant for native Dutch speech as well. The releveled model did not reveal a significant effect of Focus for native Dutch speech (β_{Focus}= -1.5, t = -1.8), indicating that the VOTs in Dutch were not longer in contrastive focus than in non-focus position.

Additionally, we see two effects that we did not observe in the *English speech* dataset. The model reveals a significant effect of Plosive, in which the /p/ had a shorter VOT than the /k/. Furthermore, if the plosive was preceded by silence, the VOT was shorter. To investigate whether the absence of these effects in the *English speech* dataset may be a power issue, we took the pre-final model, which included the residual outliers, and reduced the *Dutch speakers* dataset to the same size as the *English speech* dataset to see if the effects of Plosive and Previous Phone remained in the smaller sample. We then removed the residual outliers and found that the fixed effect for Plosive (β_{p} = -12.5, t = -2.1) and Previous Phone ($\beta_{silence}$ = -4.1, t = -2.0, β_{voiced} = -1.6, t = -0.9) remained significant.

The random structure revealed that VOT differed per Speaker and per Target Word (significant random intercepts of Speaker and Target Word) and that the effect of Speech Style varied per Speaker (significant random slope of Speech Style by Speaker).

2.4.4.3 VOT Interim Discussion

Our VOT data revealed that VOTs were longer in native English than in non-native English and they were even shorter in Dutch. This is in line with past research that native English VOT production of /p/ and /k/ is longer than those produced by native Dutch (e.g., Lisker & Abramson, 1964). Furthermore, the difference between native Dutch and non-native English VOT length indicates that when speaking non-native English, the Dutch are adapting to a certain extent and lengthening their VOT.

In general, unexpectedly, VOTs were shortened in Lombard speech. More importantly for our research question, in the *Dutch speakers* dataset, the Dutch speakers were affected differently by Lombard speech in their native Dutch and non-native English, shortening their VOT less in Dutch Lombard speech than in non-native English Lombard speech. This was not the case for the *English speech* dataset, where the effect of Lombard speech was similar for native and non-native English. Combined, this indicates that the non-native English speakers were making changes to VOT similar to natives in their non-native Lombard speech, but unlike what they do in their native speech.

Additionally, we observed that focus had an effect on the VOT data in English; words with contrastive focus having lengthened VOTs, and this was more so the case for native English speech. In contrast, for native Dutch, the VOT was not affected by focus. In terms of focus, we thus find differences between native Dutch and non-native English on the one hand, while also finding differences between native and non-native English on the other hand.

As for the control variables, for the *English speech* data, duration of the following segment was influential in lengthening VOT, in line with Hazan and colleagues' (2012) research on clear speech. Additionally, in English, VOT was shorter in following occurrences of the word. Surprisingly, we did not find a difference between the /p/ and /k/ plosives in the *English speech* data. For the control variables in the *Dutch speakers* dataset, the VOTs of /p/ were shorter than of /k/, in line with previous research (Lisker & Abramson, 1964).

2.5 Discussion

In this study, we investigated non-native Lombard speech acoustics. Non-native Lombard speech may differ from native Lombard speech because non-natives have a higher cognitive load when speaking in a non-native language (Kormos, 2006; Segalowitz, 2010) and this may be even more the case in noise. Moreover, non-native Lombard speech adaptations may show the signature of the speakers' native language. So far, only a few studies have examined non-native Lombard speech, restricting themselves to a small number of acoustic cues that are known to be modified in Lombard speech (intensity, vowel duration, F0 and F0 range; Cai et al., 2020, 2021; Mok et al., 2018; Villegas et al., 2021; Chapters 3, 4).

For the purpose of this study, we compiled the Dutch English Lombard Native Non-Native (DELNN) corpus, which is the first corpus to include non-native Lombard speech, in combination with native speech of the two languages involved. The DELNN corpus includes nine native American-English female speakers producing English plain and Lombard speech and 30 native Dutch female speakers producing native Dutch and non-native English plain and Lombard speech. These speakers read 144 English question-answer pairs, half of which were produced as plain speech and the other half as Lombard speech. Furthermore, a target word – words selected for their difficulty for Dutch speakers in English – was embedded in each answer. These target words constituted three categories: words starting with /θ/, a phoneme which does not occur in Dutch, English-Dutch cognates with schwa in prestress position in English and a full vowel in the Dutch counterpart, and words ending in voiced obstruents, which do not occur in Dutch because of final devoicing. Of note, each answer contained contrastive focus. For half of the answers, the target word received contrastive focus (late-focus condition), while for the other half, contrastive focus was earlier in the sentence (early-focus). The native Dutch speakers additionally read 96 Dutch question-answer pairs, also of which half were produced as plain speech and half as Lombard speech, and half with early-focus and half with late-focus.

Using the DELNN corpus, we examined four acoustic measures from the answers in the question-answer pairs: intensity and spectral CoG of the complete answers, durations of the target words, and VOTs of a subset of these target words. For each acoustic measure, there were two comparisons, one examining native and non-native English speakers (*English speech*) and the other examining the same non-native speakers in their non-native English and native Dutch (*Dutch speakers*).

Our analyses revealed that the non-native speakers were producing Lombard speech, adapting all four acoustic measures in noise in the same direction as the native English speakers and as in their native language, increasing intensity, spectral CoG, and word duration, and decreasing VOT compared to plain speech. These adaptations also been shown in previous studies dedicated to native speech, for intensity, spectral CoG, and word duration (e.g., Dreher & O'Neill, 1957; Lu & Cooke, 2008; Van Summers et al., 1988). For non-native Lombard speech, research has been done on intensity, vowel duration, F0, and F0 range, but this study is the first to document changes in spectral CoG, word duration, and VOT for non-native Lombard speech. Further, the current article as well as the research of Shen (2022) adds to the limited research on native Dutch Lombard speech.

In fact, our study is the first one documenting an effect of Lombard speech on VOT, either in native or in non-native speech. Our data showed shorter VOTs for the voiceless plosives /p/ and /k/ as compared to plain speech. To our knowledge, there have not been any studies investigating VOT and Lombard speech, which did not find an effect if they took word

duration into account, which we did as well by including a duration control measure. Further research with more controlled stimuli is needed to better understand why in our research, we found a decrease in VOT length in Lombard speech.

Our analyses not only showed that the non-natives adapted the four acoustic measures in noise in the same direction as the native English speakers, but also that they did so to a similar extent. This shows that non-native speakers need not adapt their speech in noise less or differently than native speakers do just because they are non-native. At least the non-native speakers we investigated in this study (native speakers of Dutch speaking English at a B2 level, as per the European Common Framework; Council of Europe (2001) adapted their speech as much as native speakers.

Since we did not observe differences between how the native and non-native English speakers adapted their speech in noise for the four measures, the comparison between native Dutch and non-native English Lombard speech may show why. It may show whether a difference is absent because the two languages adapt their speech in a similar manner in noise, or because the Dutch speakers have learned how to successfully adapt their Lombard speech in their non-native language.

When comparing native Dutch and non-native English (from the same speakers), we found differences in their Lombard speech adaptations for three of the four measures. While the speakers increased their intensity in Lombard speech in the late-focus sentences both in their non-native English and their native Dutch, they did less so in their native Dutch. With respect to the increase in spectral CoG in Lombard speech, this increase was larger in native Dutch than in these speakers' non-native English. Finally, the shortening in VOT in Lombard speech was smaller in native Dutch compared to non-native English. These differences in Lombard speech adaptations for these three measures between native Dutch and non-native English speech by the same speakers is not likely to result from differences in stimuli in the two languages. If the stimuli were influential (considering that native Dutch speech consisted of different stimuli than non-native English speech), then we would expect to see a difference in plain speech between the Dutch and English stimuli as well, rather than the differences that emerge only in Lombard speech. Combined, these results would indicate that the native Dutch speakers adapt the three acoustic characteristics differently in their Lombard speech in native Dutch and in non-native English. Future research should investigate why differences between native Dutch and non-native English emerge for certain measures and not others, as it remains unclear.

Together, the two comparisons (of the native and non-native English and of the native Dutch and non-native English) suggest that the Dutch speakers adapted their non-native English to native English, when producing Lombard speech, and were not influenced by their

native language in this respect. This may be surprising considering that research on median F0 and F0 range in non-native Lombard speech suggests that the non-native English speakers are influenced by their native Dutch (Chapters 3, 4). This may indicate that it may depend on the acoustic measure whether we see a native language influence or not: one acoustic cue of Lombard speech may be easier to adapt than another one. This calls for further research, into more acoustic measures.

While our data indicate that the Dutch speakers produced non-native English Lombard speech similarly to native English speakers, it is unclear why this the case. Dutch learners of English do not explicitly learn (for instance, at school) how to produce Lombard speech in English. Perhaps they learn this unconsciously, for instance, when watching English spoken movies. Another possibility is that some of the phonological or phonetic differences between Dutch and English trigger differences in how acoustic characteristics are adapted in Lombard speech. This calls for further research.

Future research should also extend our research to different language pairs. We chose to investigate non-native English as produced by native speakers of Dutch, because many Dutch speakers are so proficient in English that they can be expected to produce Lombard speech. Moreover, Dutch and English are very similar to each other, but also show differences, for instance, in VOT, which made it likely that differences between native and non-natives speech may be found. Of note, we did not find differences in the extent of decrease in VOT length in Lombard speech for native and non-native English. Future research could focus on language pairs that differ more substantially from each other, which may enlarge the chance that differences between native and non-native Lombard speech will be observed.

As mentioned, the native Dutch speakers always completed the English session before the Dutch session, presenting a confound. However, these sessions were completed on separate days, and we do not expect that the participants behaved differently during the second session. Additionally, the plain condition always preceded the Lombard condition. Here, we also do not expect the order of the conditions to affect the results as we started with the less demanding condition, but future research could investigate the potential effect of the order of the session.

In the sentences read by the participants, the location of contrastive focus was manipulated, and we therefore included it as a predictor for the four acoustic measures in our analyses. Target words with contrastive focus were longer than those without, in native English, non-native English, and in native Dutch, in line with past research (e.g., Cooper et al., 1985; Sityaev & House, 2003). However, the native English speakers lengthened the words with contrastive focus more so than the non-native English speakers in English, while the Dutch native speakers showed even more lengthening for words in focus in Lombard

speech, both in Dutch and in non-native English. With respect to VOT, contrastive focus lengthened VOT more in native English than in non-native English, and not in native Dutch. Further, late-focus led to a higher average sentence intensity in native and non-native English Lombard, probably because fewer words in the late-focus condition than in early-focus condition underwent post-focus compression (PFC, e.g., Xu, 2011), where material after the focus is accompanied by a decrease in intensity (e.g., Chen, 2015). Finally, while focus has been shown to affect the distribution of energy in speech (e.g., Campbell, 1995; Campbell & Beckman, 1997), data from spectral CoG of the utterance, our chosen measure of energy, did not show an effect of focus. It could be the case that spectral CoG is affected by whether there is contrastive focus in the sentence, and not so much on where the contrastive focus is located. Combined these data patterns indicate that native and non-native English speakers were implementing focus (slightly) differently, where the non-native speakers were not just implementing focus as they did in their native language.

In conclusion, this article expands upon non-native speakers' production of Lombard speech by examining four distinct acoustic measures: intensity, spectral CoG, word duration, and VOT. We did not observe differences in how native speakers of English and of Dutch adapt their English speech in noisy conditions, indicating that the non-native English are producing Lombard speech similarly to the native English. Importantly, the comparison of the native Dutch and non-native English sentences produced by the same participants nevertheless suggests that, for several acoustic measurements, the Dutch speakers adapt their speech differently in native Dutch than in non-native English. Combined, this would indicate that, when speaking English, Dutch speakers adapt their way of speaking in noisy conditions to native English.

Chapter 3:
Pitch in native and non-native Lombard speech

Abstract

Lombard speech, speech produced in noise, is typically produced with a higher fundamental frequency (F0, pitch) compared to speech in quiet. This paper examined the potential differences in native and non-native Lombard speech by analyzing median pitch in sentences with early- or late-focus produced in quiet and noise. We found an increase in pitch in late-focus sentences in noise for Dutch speakers in both English and Dutch, and for American-English speakers in English. These results show that non-native speakers produce Lombard speech, despite their higher cognitive load. For the early-focus sentences, we found a difference between the Dutch and the American-English speakers. Whereas the Dutch showed an increased F0 in noise in English and Dutch, the American-English speakers did not in English. Together, these results suggest that some acoustic characteristics of Lombard speech, such as pitch, may be language-specific, potentially resulting in the native language influencing the non-native Lombard speech.

This chapter is an edited version of:

Marcoux, K., & Ernestus, M. (2019). Pitch in native and non-native Lombard speech. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2605–2609). Melbourne, Australia: Canberra, Australia: Australasian Speech Science and Technology Association Inc.

3.1 Introduction

Many studies have documented the characteristics of speech produced in noise, Lombard speech, using native speakers. From this research, we know that Lombard speech has specific acoustic characteristics that differentiates it from speech produced in quiet. Among other features, Lombard speech is characterized by having a higher fundamental frequency (F0, pitch), a higher amplitude, and a shift in energy to higher frequencies (e.g., Castellanos, Benedí, & Casacuberta, 1996; Junqua, 1993; Lu & Cooke, 2008; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Research done to date has extensively studied native Lombard speech, primarily in English (e.g., Lu & Cooke, 2008; Van Summers et al., 1988), but also in other languages including Spanish (e.g., Castellanos et al., 1996), and French (e.g., Garnier & Henrich, 2014). However, very little research has been done with non-natives' production of Lombard speech and the resulting acoustic characteristics. This study contributes to our knowledge of Lombard speech by comparing Lombard speech produced by natives and non-natives.

There are compelling reasons to assume that there may be differences between native and non-native Lombard speech. First, the native language is known to influence the non-native language in many domains. This can be observed, for instance, in difficulties in perception and production of non-native phonemes (e.g., Best, 1995; Cutler, 2012; Flege, 1987). We therefore may observe that how non-native speakers adapt to a noisy environment reflects how they do so in their native language.

Second, we must consider the higher cognitive load that non-natives experience when speaking in their non-native language (e.g., Kormos, 2006; Segalowitz, 2010). Due to this higher cognitive load, non-natives may be less effective in adapting their speech in noise.

This study focused on differences in pitch between speech produced in quiet and in noise, as one fundamental characteristic of Lombard speech is higher pitch. Moreover, we know that different languages have both different pitch ranges as well as mean pitches. For Dutch women, a mean pitch of 191 Hz was found by van Bezooijen (cited in van Bezooijen, 1995). This mean pitch is higher for American-English (AmE) women of a similar age at 224 Hz (Stoicheff, 1981). Research on bilinguals further illustrates mean pitch differences among languages. For instance, Voigt, Jurafsky, and Sumner (2016) examined German-French and German-Italian bilinguals, finding that individuals had different mean pitches in their two languages. Finally, several studies (e.g., Rasier, & Hiligsmann, 2007) have shown that pitch patterns in the native language influence its production in the non-native language. Collectively, this research illustrates that pitch is a promising feature of Lombard speech that may differ between native and non-native speakers.

We focused on median pitch, examining AmE and Dutch speakers in their native

languages as well as Dutch speakers in their non-native English. Dutch speakers tend to show an influence of an AmE accent when speaking in English, so we chose native AmE speakers as a comparison. As there is no research to our knowledge on native Dutch Lombard speech, we do not know whether there are differences in native AmE and native Dutch in pitch in Lombard speech.

English and Dutch differ in their median pitch. For instance, the variant British English (RP) has a wider pitch range than Dutch, with lower lows and higher highs (e.g., Gussenhoven & Broeders, 1997). Importantly, RP and Dutch speakers have similar pitch accents at the sentence level in their native languages (e.g., Gussenhoven & Broeders, 1997). As a consequence, comparing median pitch is informative.

For our study, participants read sentences in quiet and in noise, which elicited native and non-native plain and Lombard speech. We manipulated the location of focus in the sentence to have early- or late-focus, expecting a median pitch difference in the two sentence types due to post-focus compression (e.g., Xu, 2011). Post-focus compression narrows and lowers the pitch range for material after the focused word.

3.2 Methods

3.2.1 Participants

Thirty native Dutch females from Radboud University, Nijmegen, the Netherlands (RU), and nine native AmE females studying abroad at RU, with an average age of 21.33 and 22.11 years, respectively, participated in the study. The Dutch participants had native Dutch speaking parents, and had completed or were completing their studies in Dutch. On average, they had an English level of B2 in the Common European Framework (Council of Europe, 2001), as indicated by their LexTALE (Lemhöfer & Broersma, 2012) scores (mean = 69.39, standard deviation = 15.76). All participants reported no hearing or vision problems, as well as no dyslexia or stuttering. The participants were given course credit or gift vouchers in exchange for their participation.

3.2.2 Speech Materials

As we wanted to examine the effect of noise on sentence median pitch and expected that median pitch is modulated by the position of focus in the sentence, we had four conditions: quiet early-focus, quiet late-focus, noise early-focus, and noise late-focus. We manipulated the location of focus in the sentence using contrastive question-answer pairs.

An example of early- (1) and late-focus (2) question-answer pairs are presented below:

1. *Did the **friends** go to the parade in Barcelona?*
   *No, the **family** went to the parade in Barcelona.*

2. *Did the family go to the **beach** in Barcelona?*
   *No, they went to the **parade** in Barcelona.*

There were 144 English and 96 similarly structured Dutch sentence pairs, half with early- and half with late-focus. These pairs were randomized within condition per language three times to create three separate master lists. The three lists were then mirrored so the pairs in the quiet condition appeared in the noise condition and vice versa, with the order of the stimuli remaining the same within condition. This resulted in six lists. The two quiet conditions always preceded the noise conditions and the order of the early- and late-focus conditions were counterbalanced. Every participant read one list.

3.2.3 Procedure

Participants recorded the 144 English question-answer pairs at their own pace while wearing Sennheiser HD 215 MKII DJ headphones in a sound attenuated room. During the quiet condition, nothing was played via the headphones, while in the noise condition, Speech-Shaped Noise at 83 dB SPL (77 dBA, as calibrated using the Brüel & Kjær Type 4153 artificial ear) was played through them using an ASUS X52J laptop. The recordings were made with a Sennheiser ME 64 or 65 microphone placed 15cm away from the participants' mouth. The microphone fed into a preamplifier and a Roland R-05 WAVE/MP3 Recorder, resulting in 44.1 kHz sampling rate with 16-bit resolution wav file.

After the English recording session, the Dutch participants completed the English LexTALE task (Lemhöfer & Broersma, 2012) which gauges English proficiency, and a demographics and language questionnaire. Within one week, the Dutch participants returned for a second session to read the 96 Dutch question-answer pairs. The first session took one hour and the second session forty-five minutes.

3.2.4 Pre-processing of the data

The audio was segmented at the sentence level. We extracted F0 values only from the answers, using Praat (Boersma & Weenink, 2018). The Praat script returned F0 values at 10 millisecond intervals. This value was -1 and excluded from analysis if the segment was

unvoiced. The pitch range was set at 75-500 Hz for all speakers.

Cleaning of the data was necessary due to pitch tracking errors, doubling and halving, and the presence of creaky voice. Doubling and halving pitch tracking errors are erroneously reported sudden jumps in the pitch by a factor of two. Prototypical creaky voice is problematic because of its irregular F0 values (e.g., Keating, Garellek, & Kreiman, 2015). By choosing 75 Hz as the minimum pitch range, speakers' creaky voice was commonly labelled between 75 and 110 Hz. We deleted doubling, halving, and creaky voice by detecting pitch jumps above or below a factor of 1.5. This meant that sudden changes in pitch as well as creaky voice were eliminated and not used to calculate the median F0 of the answer.

From these cleaned data, we calculated the minimum and median F0 value for each answer sentence. Answers with a minimum value below 110 Hz were excluded from analyses as this was an indication that creaky voice was still present. This left us with 91.75% of the original dataset (7,794 of 8,495 median F0 values which were roughly equally distributed over quiet and noise and early- and late-focus).

3.2.5 Analyses

In order to analyze the median pitch of our data using linear mixed effects models (lmers), with participant and sentence as crossed random effects, we used the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2016). First, we analyzed native versus non-native English using the native AmE and non-native English data. We then compared native and non-native speech within speaker, using the native Dutch and non-native English data.

Prior to conducting the lmers on median pitch, we removed outliers, as defined as values 2.5 standard deviations above or below the grand mean. Our fixed effects were nativeness (native, non-native), focus (early, late), and noise (quiet, noise), and the control predictor trial number. We used anovas on nested models, or AIC scores when not nested, to determine if inclusion of the effects and their interactions significantly improved the model. We tested random slopes of the fixed effects (by-subject and/or by-sentence) using the same method. For the final model, we removed data points with standardized residuals above 2.5 standard deviation units from the last model and refitted it.

3.2.6 Results

3.2.5.1 Native versus Non-Native English

As is reflected in Figures 1 and 2, the final model using data from the native AmE and the non-native English revealed a three-way interaction between noise, nativeness, and focus. We split the data by focus to better interpret it.

The late-focus model established a significant simple effect of noise (β_{noise} = 13.53; t = 7.36), with no significant effect of nativeness ($p > 0.05$) or interaction of nativeness and noise ($p > 0.05$). This indicated that the native AmE and the non-native English behaved in the same way, both groups increasing their median pitch when speaking in noise as compared to quiet. The effect of noise differed per participant and per sentence, as indicated by the random slopes.

The early-focus model revealed a significant simple effect of noise (β_{noise} = 13.23; t = 5.65), which was modulated by nativeness ($\beta_{noise\ x\ native}$ = -10.00; t = -2.05), and by a random slope of noise by participant. The non-native English were more affected by noise than the native AmE, the former having a larger increase in pitch going from quiet to noise. When the data was further split to examine the native AmE data, there was no effect of noise (β_{noise} = 3.09; t = 0.45). In contrast, the non-native English data showed an effect of noise (β_{noise} = 13.25; t = 7.51).

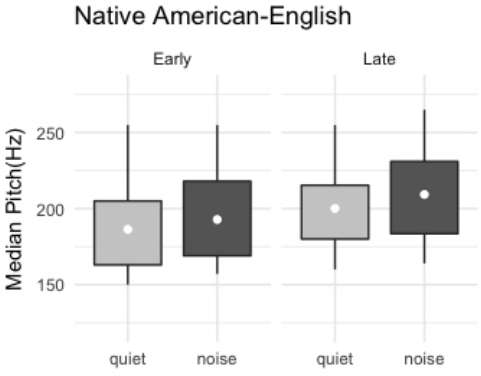

Figure 1: *Boxplot of the median pitch values of native American-English in early- and late-focus in quiet and noise. The white dots represent the means.*

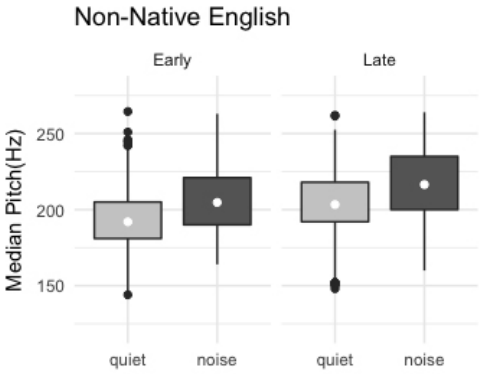

Figure 2: *Boxplot of the median pitch values of non-native English in early- and late-focus in quiet and noise. The white dots represent the means.*

3.2.5.2 Native Dutch versus Non-Native English

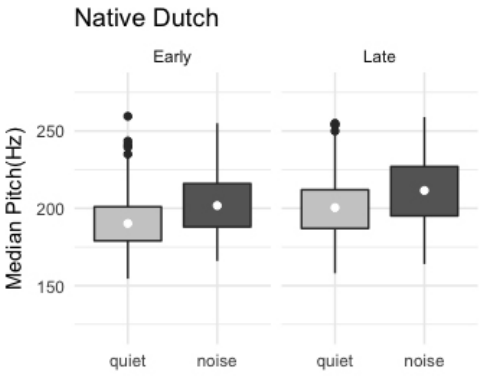

Figure 3: *Boxplots of the median pitch values of native Dutch in early- and late-focus in quiet and noise. The white dots represent the means.*

Figures 2 and 3 illustrate the findings from the final model using data from the native Dutch and the non-native English, revealing significant simple effects of noise (β_{noise} =10.42; t = 5.45), nativeness (β_{native} = -1.94; t = -3.80), focus ($\beta_{late\text{-}focus}$ = 11.15; t = 9.77), and trial number (β = 0.040; t = 4.46), with no interactions. The effect of noise and focus differed per participant as indicated by the random slopes. Our model indicates that the median pitch

increased going from quiet to noise, as well as going from early-focus to late-focus, and also that there was a small decrease in median pitch when going from non-native English to native Dutch. Additionally, over the course of the experiment itself, pitch increased with trial number.

3.3 Discussion

In our study, we compared how native Dutch speakers modulate their median pitch in Lombard speech in native Dutch and in non-native English and how native AmE speakers do so in English. We examined potential pitch differences in native and non-native Lombard speech due to non-native speakers' higher cognitive load and/or possible influences of the native language.

The comparison of the native and non-native English data showed a difference between early- and late-focus sentences. The late-focus data showed an effect of noise and no effect of nativeness, indicating that the native AmE and the non-native English speakers had a higher median pitch to the same extent in noise, an indication of Lombard speech. This showed that despite non-natives experiencing a higher cognitive load when speaking, they adapted to background noise to the same extent as native speakers. This is in line with previous research that considered Lombard speech production to be automatic, "Lombard reflex" (e.g., Van Summers et al., 1988).

While the native and non-native English speakers thus showed the same pattern in late-focus, they differed in early-focus sentences. In early-focus sentences, we saw that the non-native English showed a larger increase in pitch in noise than the native AmE. It seems that there may be a larger effect of post-focus compression in native than in non-native English.

Because of this difference between native and non-native English, we examined native Dutch data to help determine the potential influence of the native Dutch on the non-native English Lombard speech. From this comparison, we saw that the native Dutch's median pitch was slightly higher when speaking in non-native English than in Dutch, which is in line with research showing that native AmE pitch is higher (e.g., Stoicheff, 1981) than native Dutch pitch (e.g. van Bezooijen, 1995).

More importantly, we saw that the Dutch speakers had the same pattern of change going from speech in quite to noise in their native and non-native languages. In both languages, they showed an effect of noise, leading to an increase in median pitch, a characteristic of Lombard speech. They also showed an effect of focus, in which late-focus sentences had a higher median pitch than early-focus sentences, as likely explained by post-focus compression (e.g., Xu, 2011).

The same pattern thus held for the Dutch participants in native Dutch as in non-native English; an effect of noise, an effect of focus (in addition a slight effect of language). Meanwhile, there is a difference between native and non-native English. Combined, these data suggest an influence of the native language in the non-native speech, both in quiet and in noise, and consequently that there are language differences in Lombard speech.

If there are language differences in Lombard speech, we wonder how much Lombard speech differs per language, and how much of a reflex Lombard speech truly is. Further research is needed to determine whether other characteristics of Lombard speech also show the influence of the native language on non-native speech. The authors plan to analyze other acoustic measures, including pitch range and intensity to further examine the role of the native language.

Potential language specific characteristics of Lombard speech may account for recent findings on how non-native listeners perceive native Lombard speech. Native listeners understand speech presented in noise better when it was also produced in noise (Lombard speech) than when it was produced in quiet (e.g., Dreher & O'Neill, 1957; Lu & Cooke, 2008; Pittman & Wiley, 2001; Van Summers et al., 1988). This Lombard benefit is smaller for non-native listeners (e.g., Cooke & García Lecumberri, 2012). Possibly this is the case because non-native listeners do not benefit as much from Lombard characteristics that differ subtly in their native languages. Testing the perception of non-native Lombard speech using this dataset will further yield insight into non-native Lombard speech.

In conclusion, by examining pitch in native and non-native speech produced in quiet and in noise, we gain insight into potential language differences in Lombard speech. Despite experiencing a higher cognitive load, non-natives successfully produce Lombard speech in terms of increasing their pitch. Importantly, we saw a difference from native AmE speakers, indicating the influence of the native language on the non-native Lombard speech.

Chapter 4:
Differences between native and non-native Lombard speech in terms of pitch range

Abstract

Lombard speech, speech produced in noise, is acoustically different from speech produced in quiet (plain speech) in several ways, including a higher and a wider F0 range (pitch). Extensive research on native Lombard speech does not consider that non-natives experience a higher cognitive load while producing speech and that the native language may influence the non-native speech. We investigated pitch range in plain and Lombard speech in native and non-natives.

Dutch and American-English speakers read contrastive question-answer pairs in quiet and in noise in English, while the Dutch also read Dutch sentence pairs. We found that Lombard speech is characterized by a wider pitch range than plain speech, for all speakers (native English, non-native English, and native Dutch). This shows that non-natives also widen their pitch range in Lombard speech. In sentences with early-focus, we see the same increase in pitch range when going from plain to Lombard speech in native and non-native English, but a smaller increase in native Dutch. In sentences with late-focus, we see the biggest increase for the native English, followed by non-native English and then native Dutch. Together these results indicate an effect of the native language on non-native Lombard speech.

This chapter is an edited version of:

Marcoux, K., & Ernestus, M. (2019). Differences between native and non-native Lombard speech in terms of pitch range. In M. Ochmann, M. Vorländer, & J. Fels (Eds.), *Proceedings of the ICA 2019 and EAA Euroregio. 23rd International Congress on Acoustics, integrating 4th EAA Euroregio 2019* (pp. 5713–5720). Berlin, Germany: Deutsche Gesellschaft für Akustik. https://doi.org/10.18154/RWTH-CONV-239240

4.1 Introduction

Often, we find ourselves in situations surrounded by background noise; supermarkets, restaurants, and cafeterias to name a few. In these noisy conditions, we produce Lombard speech, to counter the noise, which is acoustically different from speech produced in quiet (plain speech). Among other features, Lombard speech is characterized as having a higher fundamental frequency (F0, pitch), a shift in the energy to higher frequencies, and a higher intensity (e.g., Castellanos, Benedí, & Casacuberta, 1996; Junqua, 1993; Lu & Cooke, 2008; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Some researchers have also found that Lombard speech is accompanied by a larger pitch range (e.g., Garnier & Henrich, 2014; Welby, 2006). The extensive research on Lombard speech to date has focused on native Lombard speech in various languages, including English (e.g., Lu & Cooke, 2008; Van Summers et al., 1988), French (e.g., Garnier & Henrich, 2014), and Spanish (e.g., Castellanos et al., 1996). Importantly, there has been little to no research on non-native Lombard speech especially in terms of acoustic analyses. We expand upon previous research and examine pitch range in native and non-native Lombard speech.

There are several reasons for why we may expect non-native Lombard speech to differ from native Lombard speech. First of all, extensive research has documented the influence of the native language on the non-native language. This has been observed in several domains including phonetics, speech production and perception of non-native phonemes (e.g., Best, 1995; Cutler, 2012; Flege, 1987) and prosody (e.g., van Maastricht, Krahmer, & Swerts, 2016). Thus, we may expect that pitch range in the native language may affect pitch range in the non-native language, both in speech in quiet and in noise. Second, we must consider the higher cognitive load that non-natives experience when speaking in a non-native language (e.g., Kormos, 2006; Segalowitz, 2010). Due to non-native speakers' higher cognitive load, we are unsure how and to what extent they adapt to the additive noise when producing Lombard speech in terms of pitch range.

Taking these two factors into account, the non-native speakers may change various acoustic cues to a different extent when going from plain to Lombard speech than native speakers. This is in line with an earlier study (Chapter 3), where we found differences in median pitch between Lombard speech produced by native and non-native English speakers. For sentences with early-focus, the Dutch increased their median pitch more in non-native English than the native American-English speakers, which is in line with what the Dutch do in their native Dutch (Chapter 3).

We further investigated pitch in native Dutch and native and non-native English, now focusing on pitch range. Importantly, Dutch and British-English (RP variant) differ in terms of their pitch range. Gussenhoven and Broeders (Gussenhoven & Broeders, 1997) discussed

how RP has lower lows and higher highs, indicating that RP has a wider pitch range than Dutch. Additionally, Dutch and English speakers are known to have different mean pitches. American-English females were reported to have a mean pitch of 224 Hz (Stoicheff, 1981), while similarly aged Dutch females were found to have a lower mean pitch of 191 Hz (as cited in van Bezooijen, 1995). These differences in mean pitch per language are unlikely to be due to physical differences, as Voigt, Jurafsky, and Sumner (2016) found that bilingual speakers had different mean pitches in their two languages. Dutch and English have different mean pitch and pitch range, which may affect the native and non-native Lombard speech production.

In the current study we examined pitch range in native (American-English speakers) and non-native (Dutch speakers) English for plain and Lombard speech. We also compared the Dutch when speaking their non-native (English) and their native (Dutch) language. Because the corpus on which we based our analysis also manipulated the position of focus, we took focus position into account.

4.2 Methods

4.2.1 Participants

The Dutch English Lombard Native Non-Native (DELNN) corpus was used for the present study (Chapter 2). This corpus consists of recordings from 30 native Dutch females (average age: 21.33 years) who were studying or had had completed their studies in Dutch at Radboud University, Nijmegen, the Netherlands (RU). They all had Dutch speaking parents and had an average level of B2 for English in the Common European Framework (Council of Europe, 2001), as indicated by their LexTale (Lemhöfer & Broersma, 2012) scores (M=69.39, sd = 15.76). The corpus additionally includes speech from nine native American-English females (average age: 22.11 years), who were studying abroad at RU at the time of the creation of the corpus. American-English speakers were chosen as a baseline because Dutch speakers show an influence of American-English in their speech. Participants did not report any hearing or vision problems, nor stuttering or dyslexia. All participants gave informed consent, and, in exchange for their participation, they received course credit or gift vouchers.

4.2.2 Speech Materials/Stimuli

Participants read contrastive question-answer pairs. In example (1) below we present an early-focus question-answer pair and in (2) we have a late-focus pair (words printed in bold received contrastive focus).

(1) 'Did the **friends** go to the parade in Barcelona?
No, the **family** went to the parade in Barcelona.'

(2) 'Did the family go to the **beach** in Barcelona?
No, they went to the **parade** in Barcelona.'

We had a total of 144 English sentence pairs as well as 96 Dutch pairs which were structured in a similar manner. Half of the sentences in each language were early-focus and half were late-focus.

The pairs were pseudorandomized three times by focus-condition and language creating three master lists per language. For each master list, half of the sentences in each focus condition were assigned to the quiet condition (participants produced them without background noise) and the other half to the noise condition (with background noise). This led to the four conditions: quiet early-focus, noise early-focus, quiet late-focus, and noise late-focus.

These three master lists were then mirrored, so that the sentences that appeared in quiet appeared in noise and vice-versa. This resulted in six lists per language. In all of them, the quiet conditions were followed by the noise conditions; the early- and late-focus sentences were blocked and the order of these blocks was counterbalanced across lists. Each participant was assigned one list.

4.2.3 Procedure

The recording session took place in a sound attenuated room, during which participants wore Sennheiser HD 215 MKII DJ headphones. In the quiet condition, nothing was played via the headphones. In contrast, in the noise condition, Speech-Shaped Noise was played at 83 dB SPL as calibrated using the Brüel & Kjær Type 4153 artificial ear. The participants recorded the stimuli using a Sennheiser ME 64 or 65 microphone, which fed into to a preamplifier and was connected to a Roland R-05 WAVE/MP3 Recorder. This resulted in 16-bit resolution wav files with a sampling rate of 44.1 kHz.

The recording of the English stimuli was followed by the LexTale task (Lemhöfer & Broersma, 2012) for the Dutch participants, which provides an objective measure of their English proficiency level. All participants then completed a language and background questionnaire. In total, this session took approximately one hour. Dutch participants were then asked to return within one week to record the Dutch stimuli, which took approximately an additional forty-five minutes.

4.2.4 Pre-processing of data

We segmented the recordings at the sentence level and calculated the pitch range for each of the answers. The Praat (Boersma & Weenink, 2018) script extracted the F0 values every 10 milliseconds and returned the value -1 for unvoiced segments, which was excluded from our analyses.

Before calculating the pitch range, we cleaned the data to remove pitch tracking errors (doubling and halving) as well as creaky voice. Doubling and halving are pitch tracking errors, in which the pitch suddenly appears to double or halve incorrectly. Protypical creaky voice is an issue because of its irregular F0 values (Keating, Garellek, & Kreiman, 2015). We set the pitch range to 75 to 500 Hz, which meant that most creaky voice was labeled between 75 and 110 Hz. By detecting jumps and falls above a factor of 1.5, we could delete doubling, halving, and creaky voice, cleaning the data. This is the same process we used to clean the data for the median pitch analyses (Chapter 3).

We calculated the minimum F0 per answer, as well as the 10^{th} and 90^{th} percentile F0 values from this cleaned data. The minimum F0 value was only calculated as a sanity check; we eliminated answers with a minimum value below 110 Hz, as this indicated that creaky voice was still present. We calculated our pitch range over 80% (using the 10^{th} and 90^{th} percentile). We converted the Hertz values to semitones because this relates to the human perception of the pitch range, in line with previous literature (Mennen, Schaeffler, & Docherty, 2007). For the conversion, we calculated semitones with the following equation, as was used by Kitamura, Thanavishuth, Burnham, and Luksaneeyanawin (2001):

$$\text{semitone} = 12 \log_2 (\text{maximum F0}/\text{minimum F0}) \quad (1)$$

4.2.5 Analyses

We split the analyses to focus on two comparisons; 1) native versus non-native English, to investigate the effect of native versus non-native languages and 2) native Dutch versus non-native English to examine the effect of the native language on the non-native language.

Before conducting the analyses, we eliminated outliers, which we defined as 2.5 standard deviations above or below the grand mean of the pitch range as grouped in our two questions.

Using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2016), we analyzed the pitch range with linear mixed effects models (lmers), for which we included participant and stimulus as crossed random predictors. Our fixed predictors were nativeness (native, non-native), noise (quiet, noise), and focus (early, late) as well as the control predictor trial number. We tested the random slopes of the fixed predictors, namely by-participant and by-stimulus. In order to determine if the predictors of interest, their

interactions, and the random slopes significantly improved the model, we used anovas on nested models and AIC scores on non-nested models. In the case of anovas, if the comparison was significant between the two models, we took the more complex model. When using AIC scores to compare models, we chose the model with the lower AIC score. We did not include simple effects and interactions that were not significant, unless the simple effects figured in a significant interaction. Once the final model was established, we removed the data points with standardized residuals above 2.5 standard deviation units from the previous model and refitted it, which we report in the results section below. In these models, we had non-natives, quiet, and early-focus on the intercept.

4.3 Results

4.3.1 General results

Figures 1, 2, and 3 below visualize the pitch ranges produced by the English and the Dutch speakers for sentences with early- and late-focus, in quiet and in noise. For the Dutch speakers, we see both their pitch ranges in Dutch and in English. The figures indicate that Lombard speech was always produced with a wider pitch range than plain speech.

However, the increase in pitch range when going from quiet to noise differs; it was modulated by the position of the focus in the sentence, the speaker's native language, and the language in the experiment. In the early-focus sentences, when going from quiet to noise, we see that the native and the non-native English speakers increased their pitch range approximately the same amount and that the native Dutch did this to a smaller extent. In the late-focus condition, the native English had the largest increase, the non-native English had a smaller increase, and the native Dutch had the smallest increase. These patterns are confirmed by our statistical analyses.

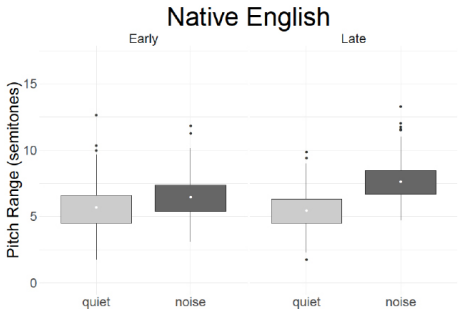

Figure 1: *A boxplot of pitch range of native English data for early- and late-focus for quiet and noise conditions. The means are represented by the white dots.*

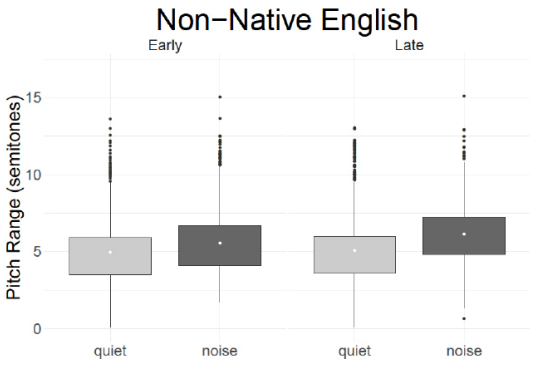

Figure 2: *A boxplot of pitch range of non-native English data for early- and late-focus for quiet and noise conditions. The means are represented by the white dots.*

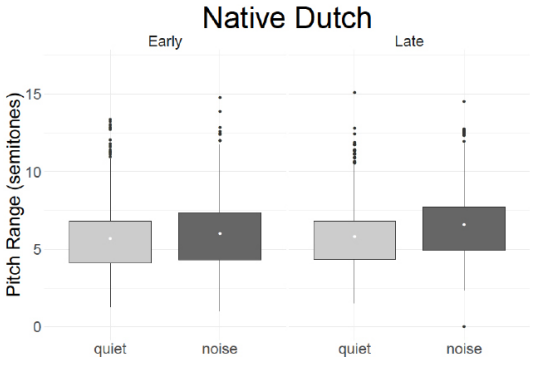

Figure 3: *A boxplot of pitch range of native Dutch data for early- and late-focus for quiet and noise conditions. The means are represented by the white dots.*

4.3.2 Native versus non-native English data

Our first statistical analysis compared the native and non-native English data, finding a three-way interaction between noise, nativeness, and focus (see Table 1 below). In order to better interpret this interaction, we split the data by focus.

Table 1: *lmer of native and non-native English Dutch data.*

	Full Model		Early-focus		Late-focus	
Fixed Effects	*β*	*t*	*β*	*t*	*β*	*t*
(Intercept)	4.96	20.00	5.15	23.08	5.02	18.56
Noise	0.51	3.74	0.55	4.94	1.07	6.66
Nativeness	0.63	1.23	-	n.s.	0.48	0.85
Focus	0.055	0.41	-	-	-	-
Noise * Nativeness	0.34	1.18	-	n.s.	1.09	3.25
Noise * Focus	0.57	9.11	-	-	-	-
Nativeness * Focus	-0.13	-0.49	-	-	-	-
Noise * Nativeness * Focus	0.64	4.36	-	-	-	-
Random Effects		*SD*		*SD*		*SD*
Stimulus (intercept)		0.37		0.39		0.34
Nativeness by Stimulus		-0.34		0.22		0.36
Speaker (intercept)		1.41		1.35		1.46
Noise by Speaker		0.57		0.64		0.78
Focus by Speaker		0.98		-		-

For the early-focus data, we found that the only statistically significant predictor of interest was noise: speakers showed a wider pitch range in Lombard speech than in plain speech. The effect of noise differed per speaker, as indicated by the random slope of noise by participant. The simple effect of nativeness and the interaction of noise with nativeness were not found to be significant (*p's* > 0.05), but there was a random slope of nativeness per stimulus. We thus did not find a general difference between the native and non-native English speakers in terms of pitch range increase in going from quiet to noise, indicating that they increased to a similar extent.

In the late-focus data, there was a significant simple effect of noise and a significant interaction of noise with nativeness. The random slope of noise by participant indicates that the effect of noise differed per speaker. The simple effect of nativeness was not significant (*p* > .05), but there was a random slope of nativeness per stimulus. Together these results indicate that the native English speakers increased their pitch range more in noise than the non-native English speakers, showing a difference between the native and non-native speakers in late-focus. This differential pattern for the two types of focus sentences (early versus late focus) explains the three-way interaction in the parent model (Table 1).

4.3.3 Non-native English versus native Dutch data

While we compared non-native English with native English in our first analysis, here we compare non-native English with native Dutch (see Table 2 below). This second analysis is thus a within-subject analysis.

We found a simple effect of nativeness as well as an interaction of nativeness with noise. Together, these effects illustrate that the Dutch speakers showed a wider pitch range in Dutch than in non-native English, especially in plain speech. In addition, we see a significant simple effect of noise and interaction of noise with focus, showing that the Dutch increased their pitch range in noise, and even more so in late-focus sentences. The effect of noise differed per participant as indicated by the random slope.

Table 2: *lmer of non-native English and native Dutch data*

Fixed Effects	*β*	*t*
(Intercept)	4.79	18.15
Noise	0.32	2.68
Nativeness	0.78	12.37
Focus	0.081	1.33
Trial	0.004	5.37
Noise * Nativeness	-0.24	-4.16
Noise * Focus	0.49	9.30
Random Effects		*SD*
Stimulus (intercept)		0.37
Speaker (intercept)		1.41
Noise by Speaker		0.57

4.4 Discussion

In this study, we examined pitch range (10^{th} to 90^{th} percentile) in plain and Lombard speech in native and non-native speakers. We investigated whether there may be differences between native and non-native speech because of non-natives' higher cognitive load and/or the influence of their native language.

One of the major findings, replicating earlier research (e.g., Garnier & Henrich, 2014; Welby, 2006), was the general increase in pitch range when going from quiet to noise. Notably, in our research, we also find that non-native speakers increase their pitch range when producing Lombard speech. This finding for the non-natives supports an earlier study

(Chapter 3), which also showed that non-natives adapt their speech in noisy environments (w.r.t. median pitch), despite their increased cognitive load.

The amount of pitch range increase varied with the position of the contrastive focus in the sentence, the speaker's native language, and the language spoken. The late-focus sentences show a clear difference between native and non-native English speakers: the native English speakers showed a larger increase in pitch range when going from plain speech to Lombard speech than the non-native English speakers. In native Dutch, the Dutch speakers showed a smaller increase. These data thus hint to an effect of non-native speakers' native language on how they produce Lombard speech: When they do not substantially increase their pitch range in their native language (Dutch), they also do not tend to do so in their non-native language (English). However, the late-focus data also showed that non-native speakers adapted their Lombard speech to the non-native language, as the Dutch speakers showed a bigger increase in pitch range in non-native English than in native Dutch, making them more similar to the native English speakers.

The early-focus sentences showed a difference between native English and native Dutch, with the native Dutch showing a smaller increase in pitch range when going from plain speech to Lombard speech. On the other hand, we did not find a difference between native and non-native English in the effect of noise on pitch range in early-focus sentences. This indicates that the non-native English speakers appear to better adapt to English for early-focus than for the late-focus sentences.

The question now arises why we found nuanced differences between early- and late-focus sentences. The answer may lie in the structure of the sentences. Both the early- and late-focus sentences formed the answers to the question-answer pairs, and started with 'No', which is likely to receive focus. The contrastive focus of the sentence then came right after the 'No'(in early-focus answers) or later in the answer (in late-focus answers). In general, when a word is in focus, it will receive higher pitch (Xu & Xu, 2005). We speculate that when the focus came later in the sentence (late-focus), the participant was able to better accent the focus word and have a higher pitch maximum, expanding the pitch range of the sentence, than when the focus was right after the accented 'No' (early-focus). This may explain, first, why we found a bigger increase in pitch range when going from plain to Lombard speech in native English than in non-native English for the late-focus sentences, but not for the early-focus sentences. Second, it may explain why there is a larger increase in pitch range for late than for early focus sentences in native Dutch and non-native English.

Overall, we found that the Dutch showed a wider pitch range in native Dutch than in non-native English. This may be unexpected because native Dutch is claimed to have a narrower pitch range than English RP (Gussenhoven & Broeders, 1997). However, in our

data, we found that, in plain speech, the pitch range is, unexpectedly, smaller in native English than in native Dutch. This may be case because we investigated contrastive question-answer pairs instead of declarative sentences.

4.5 Conclusions

We examined pitch range in native Dutch and native and non-native English speakers, observing an increase in pitch range in going from plain to Lombard speech. This is in line with our earlier finding that non-native speakers adapt their median pitch when producing Lombard speech (Chapter 3). Furthermore, we found that the non-native speakers adapted their increase in pitch range to the non-native language, more or less completely for early focus sentences, but to a less extent for the late focus sentences. The late focus sentences thus indicate that non-native Lombard speech may show properties of the speakers' native language.

Chapter 5:

The Lombard intelligibility benefit of native and non-native speech for native and non-native listeners

Abstract

Speech produced in noise (Lombard speech) is more intelligible than speech produced in quiet (plain speech). Previous research on the Lombard intelligibility benefit focused almost entirely on how *native* speakers produce and perceive Lombard speech. In this study, we investigate the size of the Lombard intelligibility benefit of both native (American-English) and non-native (native Dutch) English for native and non-native listeners (Dutch and Spanish). We used a glimpsing metric to measure the energetic masking potential of speech, which predicted that both native and non-native Lombard speech could withstand greater amounts of masking to a similar extent, compared to plain speech. In an intelligibility experiment, native English, Spanish, and Dutch listeners listened to the same words, mixed with noise. While the non-native listeners appeared to benefit more from Lombard speech than the native listeners did, each listener group experienced a similar benefit for native and non-native Lombard speech. Energetic masking, as captured by the glimpsing metric, only accounted for part of the Lombard benefit, indicating that the Lombard intelligibility benefit does not only result from a shift in spectral distribution. Despite subtle native language influences on non-native Lombard speech, both native and non-native speech provides a Lombard benefit.

This chapter is an edited version of:

Marcoux, K., Cooke, M., Tucker, B. V., & Ernestus, M. (2022). The Lombard intelligibility benefit of native and non-native speech for native and non-native listeners. *Speech Communication,* 136, 53-62. https://doi.org/10.1016/j.specom.2021.11.007

5.1 Introduction

Whether at grocery stores, restaurants, or cafes, on a daily basis we hear background noise and speak in it, producing Lombard speech (Lombard, 1911). Lombard speech is acoustically different from plain speech, that is speech produced in quiet. These acoustic modifications allow Lombard speech to be better understood in noise compared to plain speech, providing a Lombard intelligibility benefit (e.g., Dreher & O'Neill, 1957; Pittman & Wiley, 2001). Research to date has heavily focused on native speakers producing Lombard speech as well as native listeners. Considering that non-native speakers are influenced by their native language when speaking, the question arises as to whether non-native Lombard speech is produced differently and in turn how this affects the size of the Lombard benefit for non-native speech. This study is the first to investigate the perception of non-native Lombard speech.

Past research has predominantly examined native speakers' production of Lombard speech in English (e.g., Dreher & O'Neill, 1957; Pisoni, Bernacki, Nusbaum, & Yuchtman, 1985; Pittman & Wiley, 2001; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988), as well as several other languages such as French (e.g., Garnier & Henrich, 2014), Spanish (e.g., Castellanos, Benedí, & Casacuberta, 1996), and Dutch (e.g., Bosker & Cooke, 2020). These studies have established that native Lombard speech involves acoustic modifications that make it distinct from plain speech. These include but are not limited to: an increase in fundamental frequency (F0), a wider F0 range, an increase in intensity, and a shift in energy to higher frequencies (for a review see: e.g., Cooke, King, Garnier, & Aubanel, 2014).

Only a handful of studies have investigated Lombard speech produced by non-native speakers. By investigating non-native Lombard speech in combination with native Lombard speech, we can better understand Lombard speech itself. Analyzing non-native Lombard speech will reveal the potential influence of the native language on the non-native Lombard speech, leading to insights as to whether Lombard speech is language general, or whether there may be some aspects that are more language specific.

Two studies that investigated the acoustics of non-native Lombard speech (Chapters 3, 4) examined native American-English speakers and native Dutch speakers in addition to native Dutch speakers producing non-native English plain and Lombard speech. Their research used the Dutch English Lombard Native Non-Native corpus (DELNN; Chapters 2, 3). Chapter 3 found that the Dutch non-native speakers of English increased their median F0 for Lombard speech compared to plain speech, as is characteristic of native Lombard speech. They also found subtle native (Dutch) language influences on the non-native Lombard speech production in terms of the amount of F0 increase in the different conditions. The stimuli consisted of question-answer pairs, where the word with contrastive focus was early in the answer (early-focus condition) or late in the answer (late-focus condition). This meant that

the material that underwent post-focus compression (a narrowing and lowering of the F0 range for the post-focus stimuli; see e.g., Xu, 2011) differed in the two conditions. While the two groups increased their F0 to a similar extent for Lombard speech in the late-focus condition, the non-native speakers had a larger increase in F0 (average of 12.7 Hz increase) than the native English speakers (average of 6.5 Hz increase) for the early-focus condition, reflecting the larger increase in Dutch (average of 11.6 Hz increase for early-focus).

Chapter 4 also examined the increase in F0 range in Lombard speech in the natives and non-natives, finding an overall increase in F0 range in both the native and non-native speakers. Again a difference was found between the native and non-native English speakers, with the native speakers having a larger F0 range increase than the non-natives in the late-focus condition. The native Dutch had the smallest increase in Dutch, indicating that the non-native speakers were being influenced by their native Dutch. These results were not driven by individual speakers. These studies on non-native Lombard speech production, although limited to median F0 and F0 range, suggest that non-native speakers are adapting their speech in noise in ways that are characteristic of native Lombard speech, while showing faint influences of the native language.

Three other studies have investigated non-native Lombard speech in two more languages. Villegas, Perkins, and Wilson (2021) showed that Japanese speakers produced louder speech (as measured by sound pressure level) both in native Japanese and in non-native English Lombard speech relative to plain speech. Cai, Yin, and Zhang (2020) found the same with intensity for Chinese-English late-bilinguals (first language – L1 – Chinese, second language – L2 – English). Mok, Li, Luo, and Li (2018) investigated mean intensity, mean F0, and durations of vowels in L1 Mandarin and L2 English speech. They found that compared to L1 Mandarin plain speech, Lombard speech had higher mean intensity and longer durations, and for two of the three tones studied, higher mean F0. For the L2 English speech, they also found higher intensity and longer durations for Lombard speech compared to plain speech. However, the mean F0 was lower for L2 English Lombard speech in comparison to plain speech. In combination, these studies indicate that non-native speakers may produce Lombard speech, but may apply different modifications than native speakers do.

Past research with native speech has shown that the acoustic characteristics of Lombard speech provide an intelligibility benefit, resulting in Lombard speech being better understood in noise compared to plain speech presented in noise, that is, a Lombard intelligibility benefit (e.g., Dreher & O'Neill, 1957; Pittman & Wiley, 2001). In examining the Lombard benefit, the speech (plain, Lombard) is mixed with noise, and the intelligibility of the masked stimuli is measured. The signal-to-noise-ratio (SNR) is fixed at the same level for the plain and Lombard speech, allowing for a comparison of the intelligibility between the two speech styles. The Lombard benefit has been shown to be influenced by various factors, including

the type and the intensity of the noise used to elicit the Lombard speech, as well as the noise used as a masker and the SNR of the masked stimuli (e.g., Lu & Cooke, 2008; Van Summers et al., 1988).

All studies on the Lombard benefit have investigated native speech. While the vast majority has also focused on native listeners, a couple of studies have examined both native and non-native listeners. One such study was conducted by Junqua (1993). He did not find a Lombard benefit for native nor for non-native listeners.

Another study involving non-native listeners was performed by Cooke and García Lecumberri (2012), who found that non-native listeners show a Lombard benefit for native speech. In line with previous research with natives (e.g., Van Summers et al., 1988), Cooke and García Lecumberri (2012) reported that for non-native listeners, the size of the benefit also depends on the SNR level of the stimuli used in the intelligibility experiment, with higher noise levels eliciting a larger Lombard benefit. Also in agreement with past research on native listeners (Lu & Cooke, 2008), the level of noise used to elicit the Lombard speech affected the size of the Lombard benefit. The Lombard benefit for the non-native listeners, however, appears to be smaller than for native listeners who were tested with the same stimuli in another experiment (Lu & Cooke, 2008). With Lombard speech produced in 82 dB SPL of noise and plain and Lombard speech tested at an SNR of -9 dB, the native listeners increased their intelligibility 22% points when going from plain to Lombard speech (Lu & Cooke, 2008). In comparison, under the same conditions, non-native listeners increased their intelligibility by 15% points (Cooke & García Lecumberri, 2012). Together, these findings suggest that the Lombard benefit for non-native listeners is influenced by various factors similarly to native listeners (SNR levels, noise levels in producing Lombard speech, etc.) and that, although non-native listeners experience the Lombard benefit, they may do so to a smaller extent than native listeners.

One possible explanation for the smaller Lombard intelligibility benefit for non-native listeners is that in general non-native listeners can be more adversely affected than natives by noise in word recognition tasks (for a review of the literature see: García Lecumberri, Cooke, & Cutler, 2010; Scharenborg & van Os, 2019). Being more adversely affected by noise could mean that the non-native listeners may need to dedicate more cognitive resources to word recognition, and in turn may not be able to take full advantage of the Lombard cues that contribute to the Lombard benefit. Alternatively, there may be language specific modifications in Lombard speech, which the non-native listeners may not take full advantage of.

The few studies investigating the Lombard benefit for non-native listeners only examined one non-native listener group each (Cooke & García Lecumberri, 2012; Junqua, 1993). The benefit may, however, vary depending on the speaker's and the listener's native language.

In examining the intelligibility of plain speech, Bent and Bradlow (2003) found a "matched interlanguage speech intelligibility benefit". That is, a non-native listener understands a non-native speaker with whom they share the same native languages as well as a native speaker. Other studies have not found such straightforward results. For instance, Stibbard and Lee (2006) only found weak evidence for the matched interlanguage speech intelligibility benefit, and Major, Fitzmaurice, Bunta, and Balasubramanian (2002) only found the effect for one of several listener groups. Bent and Bradlow (2003) further found a "mismatched interlanguage speech intelligibility benefit," with non-native listeners benefitting from non-native speakers independently of whether they share their native language. Other research has not replicated this mismatched interlanguage speech intelligibility benefit (e.g., Stibbard and Lee, 2006). Therefore, it is still an open question whether and how a speaker's intelligibility is co-determined by the exact combination of the speaker's and listener's native languages.

In our study, we investigated the Lombard benefit of non-native speech, taking into account that this benefit may depend on the combination of the speaker's and listener's native languages. The materials (English target words) were taken from the DELNN corpus (Chapters 2, 3), which, as mentioned above, contains English speech from native (American-English) and non-native (native Dutch) English speakers as well as native Dutch speech. Lombard speech has been documented for both native English and native Dutch and the Lombard speech in the two languages show similarities (e.g., English: Bosker & Cooke, 2018; Lu & Cooke, 2008; Pisoni et al., 1985; Van Summers et al., 1988; e.g., Dutch: Bosker & Cooke, 2020). Additionally, the acoustics of English Lombard speech produced by Dutch natives have been briefly studied, and it appears that their non-native English Lombard speech is very similar to native English Lombard speech, albeit perhaps with some native language influence (Chapters 3, 4).

We first analyzed the speech signal itself, investigating the masking potential of English target words using the high-energy glimpsing proportion metric (HEGP, Tang & Cooke, 2016). The HEGP metric is an extension of the glimpse proportion metric (GP). GP measures the proportion of spectro-temporal regions (glimpses) in speech tokens where the energy is greater for the speech than for the noise (Cooke, 2006). Lu and Cooke (2008) found that Lombard speech has higher GPs than plain speech and furthermore that the GPs correlate with human intelligibility. The HEGP metric extends GP by examining each frequency band separately and selecting only those glimpses that additionally have an energy that exceeds the average speech-plus-noise energy in that band. The extension of GPs by HEGPs results in an improved correlation with intelligibility scores (Tang & Cooke, 2016). HEGPs range in value from 0 to 1, which ideally maps on to the range from 0 to 100% intelligibility, although in practice the mapping depends on the type of speech material and the masker (see Figure 3 in Tang & Cooke, 2016). This acoustic measure is independent of the listener's native language

and therefore provides language-independent intelligibility information on native versus non-native Lombard speech. As there may be native language influences on non-native plain and Lombard speech, we may expect differences in HEGPs between native and non-native speech, in addition to differences between plain and Lombard speech.

We then tested the same material in an intelligibility experiment with native and non-native listeners. Listeners were asked to identify the English target words, produced by native and non-native English speakers in plain and in Lombard speech, mixed with noise. By having one group of native and two groups of non-native listeners, we can investigate how the speaker's and listener's native languages contribute to a Lombard intelligibility benefit. Canadian listeners served as our native cohort. One non-native listener group consisted of native Dutch individuals, chosen because they shared the native language with the non-native speakers. Our other non-native listener group consisted of native Spanish individuals, chosen as they did not share the native language with any of the speakers. The two non-native listener groups allowed us to examine the matched and mismatched interlanguage speech intelligibility benefit (Bent and Bradlow, 2003).

Listeners' responses were initially analyzed independently from HEGPs. Subsequently, we performed an additional analysis on the listeners' responses in which HEGP predictions were incorporated, in order to clarify whether any seemingly language-dependent differences might be ascribed to the ability to withstand energetic masking.

5.2 Speech Materials

5.2.1 Speakers

The speech materials for this study were taken from the DELNN corpus (Chapters 2, 3). Of the nine native (American-English) and thirty non-native (native Dutch) female English speakers in the corpus, we selected eight native and eight non-native speakers. One native speaker did not agree to have her recordings used online, leaving us with the needed eight. We selected the eight mid-accented non-native speakers from the 23 who agreed to have their recordings used online based on the results from an overall accentedness rating experiment reported in Chapter 6. In the accentedness experiment, six native American-English listeners rated six sentences per non-native speaker, using a 7-point Likert scale (1 "native-like" to 7 "very strong foreign accent"). These native listeners also rated two native speakers, which resulted in averages of 1.00 and 1.03, respectively, for each of the speakers, confirming their ability to identify native speech. After averaging the ratings per non-native speaker, the eight non-native speakers who were closest to the median rating of 4.7 were selected. Their average accentedness ratings for these eight speakers ranged from 4.4 to 5.0. Further, these selected non-native speakers had an average LexTALE (Lemhöfer & Broersma, 2012) score of 64.0

(*sd* = 9.2), corresponding to a B1 level in the Common European Framework (Council of Europe, 2001). We assume these selected speakers to be normal representations of Dutch natives, who typically start learning English at the age of ten or eleven years. Dutch natives are constantly exposed to English via English movies and series, as most entertainment is not dubbed.

5.2.2 Stimuli

For the DELNN corpus, speakers read question-answer pairs at their own pace. The 96 target words were taken from the 72 early-focus sentences, where the word with contrastive focus came early in the sentence (for further details see Chapters 2, 3). In some cases, multiple target words were taken from one sentence. The majority of target words were produced as nouns (e.g., *table, gloves, city*), while some were produced as verbs but could function as nouns (e.g., *left*, *likes*, *move*). Their frequencies of occurrence ranged from 0.7 to 1958.6 in a million (*M* = 145.8, *sd* = 255.3) as reported in SUBTLEX US (Brysbaert & New, 2009). This large range in frequency of occurrence is a result of the design of the corpus, which limited us in the selection of target words. The target words always came at some point after the word with contrastive focus (anywhere from immediately after the word with contrastive focus to the last word in the sentence); thus they all underwent post-focus compression, a narrowing and lowering of the F0 range (e.g., Xu, 2011). An example question-answer pair is included in (1). Speakers were asked to place emphasis on the words in bold, resulting in contrastive focus in the answers (*Paul* in this example). From the answer in (1), we extracted the word *café*, to be used in the HEGP analysis and intelligibility experiment. See Appendix 1 for all target words used.

(1) Did **Simon** meet his professor at the café to talk?
No **Paul** met his professor at the café to talk.

A total of 662 target word tokens were chosen from the DELNN corpus, which differed from each other in the combination of speech style (plain, Lombard), speaker, and word. Not all speakers and words contributed equally to the stimuli because not all eight native and eight non-native speakers produced all 96 target words both in plain and in Lombard speech in the DELNN corpus. Each speaker contributed between 35 and 48 target word tokens. Each target word was produced the same number of times as plain and as Lombard speech. Therefore, 331 of the target words were produced as plain speech and the other 331 as Lombard speech (this ranged from two to four productions of plain / Lombard speech per target word with an average of 3.4).

The corpus was phonemically transcribed and segmented at the word level using the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017). The first author examined all oscillograms and spectrograms and improved the segmentation if needed, before the target word tokens were extracted at the zero-crossing boundaries.

5.2.3 Procedure

The speakers in the DELNN corpus completed the self-paced reading task wearing a pair of Sennheiser HD 215 MKII DJ over ear headphones. Nothing was played via the headphones for the plain speech. In contrast, to elicit Lombard speech, participants heard speech shaped noise (SSN) at 83 dB SPL (calibrated using a Brüel & Kjær Type 4153 artificial ear) through the headphones. Wearing headphones during the plain speech condition may have caused speakers to produce some amount of Lombard speech because their own voice was attenuated. We wanted the Lombard condition to purely reflect the effect of noise rather than noise in combination with headphone noise attenuation.

The SSN was generated by passing random noise through a filter whose spectrum matched the average spectrum of male and female voices. The average spectrum was created from the recordings of 10 male and 10 female adults reading a phonetically balanced text that was approximately two minutes long. The resulting SSN file had a sampling frequency of 44.1k Hz and was a single-channel WAV file, as is the case with all speech and noise materials used.

5.3 High Energy Glimpse Proportion Analysis

5.3.1 Procedure

HEGPs were calculated at a global SNR of -1 dB, as this was the SNR designated for the intelligibility experiment (Section 5.4). Each token was constructed by mixing the speech signal with a randomly-chosen fragment of the SSN masker used to elicit Lombard speech in the DELNN corpus. The speech and masker signals making up each noisy token were independently passed through a gammatone filterbank consisting of 55 filters with center frequencies in the range of 50-8000 Hz. The instantaneous Hilbert envelope at the output of each filter was subsequently smoothed with a first-order leaky integrator (8ms time constant) and downsampled to 100 Hz (i.e. 10 ms frames) for glimpse calculation. Candidate glimpses were then defined as those 10 ms time-frequency regions where the energy in the resulting spectro-temporal representation of the speech exceeded that of the masker (i.e., a local SNR of 0 dB was employed). Subsequently, in each frequency band, only those candidates occupying time regions where the speech-plus-noise mixture energy exceeded the mean energy of the

mixture in that band were retained. Details of the HEGP calculation are provided in Tang and Cooke (2016).

5.3.2 Analyses

HEGPs were transformed into logits using the equation: ln (proportion/1-porportion) (Jaeger, 2008) and analyzed using linear mixed effects models (lmers) from the *lme4* package (version 1.1.21) (Bates, Mächler, Bolker, & Walker, 2015) in R (version 3.5.1) (R Core Team, 2016). Visualizations were made using the *ggplot2* package (version 3.2.1) (Wickham, 2016). The predictors of interest were Speech Style (plain, Lombard) and Speaker Nativeness (native, non-native). Speaker and Target Word were crossed-random intercepts and the significant predictors of interest were also tested as random slopes.

In the analysis, the Nelder-Mead optimizer was used as it provided the most robust convergence results. Outliers were defined as data points more than 2.5 standard deviations away from the grand mean, and we removed 18 such outliers prior to modeling. We started with all the predictors and interactions of interest and did a backwards fitting procedure for the fixed effects structure, and then a forward fitting procedure for the random slopes. This meant that we started by including an interaction between Speech Style and Speaker Nativeness (our predictors of interest) and then removed the interaction since it was not significant ($t < 1.96$) (backwards fitting procedure). The simple effects of the predictors of interest were confirmed as significant via the summary() function ($t > 1.96$). Once the fixed effects structure was established, the significant fixed predictors were tested as random slopes as well as the interaction and anova() was used to determine if the addition improved the model (forward fitting procedure). If the model resulted in a convergence warning, we did not proceed with that model as it indicated that model was too complex given the dataset. As a final step, we took the previous model and removed the data points resulting in absolute standardized residuals exceeding 2.5 and refitted the model, which is reported below. The significance of the fixed effects of this final model were confirmed via the function summary() and via Anova(), from the *car* package (version 3.0.6) (Fox and Weisberg, 2019), which computes Type II Wald chi-square tests. For the model reported below, the levels plain speech (predictor Speech Style) and non-native speakers (predictor Speaker Nativeness) are on the intercept.

5.3.3 Results

The statistical model revealed a significant simple effect of Speech Style as well as of Speaker Nativeness (Table 1). Lombard speech had higher HEGPs than plain speech and the native speakers had higher HEGPs than the non-natives (Figure 1). The lack of a significant interaction between Speaker Nativeness and Speech Style (β= 0.0. t = -1.5) suggests that the

difference in resistance to masking between plain and Lombard speech was similar for the native and non-native speech. The random effects structure revealed that HEGPs differed per Target Word, per Speaker, and that the effect of Speech Style varied per Speaker and per Target Word.

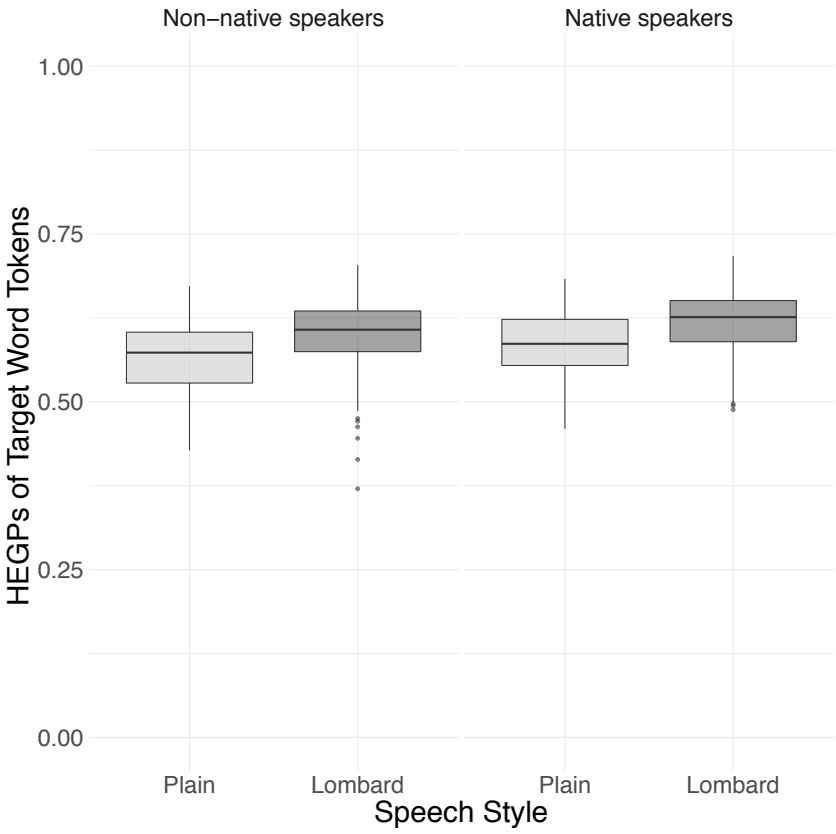

Figure 1: *HEGPs of target word tokens produced in plain and Lombard speech split by speaker nativeness. The data in the graph are the raw data (pre-logit transformation), including outliers, which were excluded from the analyses. The box represents the lower and upper quartiles, with the horizontal line within the box indicating the median. The dots indicate potential outliers while the lines extend to the minimum and maximum, excluding potential outliers.*

Table 1: *lmer model of the logit HEGPs. Plain speech and non-native speaker are on the intercept. After each predictor, we indicate which level is being contrasted with the intercept. The β indicates the size of the difference between the intercept level and the contrasted level.*

Fixed effects:	*β*	*t-value*
Intercept	0.3	9.4
Speech Style: Lombard	0.1	6.3
Speaker Nativeness: Native	0.1	2.1
Random Effects		*SD*
Target word (Intercept)		0.1
Speech Style by Target Word		0.1
Speaker (Intercept)		0.1
Speech Style by Speaker		0.1
Residual		0.1

5.4 Intelligibility Experiment

5.4.1 Methods

5.4.1.1 Participants

Our native participants were 42 Canadians (32 females) with an average age of 21.5 years (*M*) and a standard deviation (*sd*) of 3.8 years. These participants were born and raised only in English by native English-speaking parents in Canada. They had no knowledge of Dutch or German as well as no contact with Dutch or German speakers. No participant had spent more than three months in non-English speaking countries. Additionally, participants reported no hearing loss or reading problems. These participants were recruited at the University of Alberta, in Edmonton, Canada.

Our non-native participants were 47 native speakers of Spanish or Spanish/Basque bilinguals, hereafter referred to as Spanish (39 females; M = 20.2 years, sd = 2.4 years) and 46 native Dutch individuals (36 females; M = 21.7 years, sd = 2.7 years). The Spanish participants were students at the Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU) in Vitoria, Spain, while the Dutch participants were students at Radboud University, Nijmegen, The Netherlands. The Spanish participants were enrolled in English Philology at the Faculty of Arts, while the Dutch participants came from different majors, with the additional requirements that they did not study linguistics and that more than half of their classes were not in English. These different study requirements for the two non-

native groups were in place so that they had similar levels of English proficiency. On average, both the Spanish and Dutch participants had a B2 level of English in the Common European Framework (Council of Europe, 2001) as indicated by their LexTALE scores; M = 69.6, sd = 8.5 and M = 66.5, sd = 14.6, respectively (Lemhöfer & Broersma, 2012). We did not find a difference between the Dutch and Spanish participants' English proficiency as indicated by the LexTALE scores (independent non-parametric two samples Wilcoxon rank test was conducted as the Dutch data was not normally distributed, W = 862, p = 0.09). None of the non-native participants reported any hearing loss or reading problems and had not been in English-speaking countries for more than two months. Moreover, the Dutch participants were not speakers from the DELNN corpus (Chapters 2, 3). All participants gave informed consent and were compensated financially or with course credit for their participation in the experiment.

5.4.1.2 Speech Materials

The same speech materials as in the HEGP analysis were used for the intelligibility study. The 662 target word tokens were mixed with a random section of the noise from the same SSN file that was used to calculate the HEGPs. For the intelligibility experiment, a 30 millisecond ramp of noise preceded and followed the tokens. An SNR of -1 dB was chosen on the basis of a pilot which used Dutch non-native listeners of English. These pilot participants did not partake in the intelligibility experiment but had a similar background as the participants that did. Pilot participants were tested on a subset of the tokens using SNRs ranging from -2 to +4 dB. An SNR of -1 dB ensured that the performance on the easiest condition (native speakers, Lombard speech) was not at ceiling while the most difficult condition (non-native speaker, plain speech) was not at floor.

The experiment also included one filler word per speaker (see below), which were not included in the analyses. These filler words were included so the listener could adapt to each speaker. These fillers also came from the DELNN corpus and were also nouns.

5.4.1.3 Experimental lists

Each of the 12 experimental lists contained 192 trials of interest, made up of 96 distinct target words, produced twice, once by a native and once by a non-native speaker. Within each list, eight speakers (four native, four non-native) produced 24 target words each. Of the target word tokens, half were plain and half were Lombard speech. The lists were blocked by speaker and the first item of each block was a filler. The fillers were included at the beginning of each speaker block so the listeners could adapt to the individual speakers. The lists were pseudorandomized so that no more than two native or two non-native speakers followed each

other and that the second instance of the target word was at least 10 trials after the first.

To create the 12 experimental lists, we started with one list containing eight speakers – four native and four non-native. Speakers were grouped into pairs, each speaker had a corresponding speaker that produced the same target words in the other speech style (plain and Lombard). A second list was created by taking four of the speakers from the first list and adding four different speakers to it. From the two lists, we created two mirror lists, which contained the other speakers from each pair (eight speakers), and what was plain speech was now Lombard speech and vice versa. The order of speakers in these four lists and the words in each of these speaker blocks was randomized with the restrictions described above, resulting in the final 12 experimental lists. Each participant completed three practice trials (not target words) by one native and two non-native speakers (selected from our 16 speakers) followed by an experimental list.

5.4.1.4 Procedure

In the intelligibility experiment, listeners heard isolated words in plain and Lombard speech styles produced by native and non-native speakers mixed with SSN. For each trial, participants heard the stimulus while seeing a blank screen and, after 0.1 second, they were instructed to "Write down the word you heard:". If participants attempted to continue the experiment without a response, a message appeared on the screen asking them to "Please fill in the blank". The computers in the three countries had autocorrect activated, but there were misspellings that the auto-correct did not catch, such as "fundation" for "foundation". The experiment included three self-paced breaks, each after two speaker blocks.

In addition to the intelligibility experiment itself, the Dutch and Spanish participants completed the aforementioned LexTALE task (Lemhöfer & Broersma, 2012). For the task, participants needed to indicate whether the 60 test items presented orthographically were English words or non-words. Of the 60 test items, 40 were words and 20 were non-words. This task is used to estimate general English proficiency as per the levels in the Common European Framework (Council of Europe, 2001).

The Spanish participants completed the intelligibility experiment on Mac Mini computers using MacOS Sierra and Sennheiser HD 380 pro headphones. Up to four participants completed the experiment simultaneously in separate alcoves in a sound attenuated room. The Dutch participants completed the same experiment on Dell Latitude 5590 laptops using Windows 10 and Sennheiser HD 215 MKII DJ headphones. The Canadian participants completed the experiment on Dell Optiplex 3020 computers running Windows 7 and wore MB Quart QP 805 DEMO headphones. For both the Dutch and Canadian participants, up to two participants were run at once, each in their own sound attenuated booth.

The participants heard the stimuli at an average fixed volume of 71 dBA. In Spain and The Netherlands, the volume coming from the headphones was calibrated on a subset of concatenated stimuli using the Brüel & Kjaer artificial ear type 4153 and in Canada it was calibrated using the EXTECH Instruments 407750 Digital Sound Level Meter with RS232 and Sound Level Calibrator 407744.

5.4.2 Analyses

To analyze intelligibility, participants' responses were cleaned of spurious characters such as "." and "\". Further, when multiple worded or multiple answers were given, only the first word was considered (e.g., only "gang" was considered in "gang or game"). Answers were coded as correct (1) or incorrect (0). An answer was only considered correct if there were no misspellings. Correctly spelled homonyms of the target word, such as "weak" for "week" were also considered correct.

Intelligibility was analyzed with generalized linear mixed effects models (glmers) with the binomial link function. The number of maximal iterations was increased to 100,000 in "bobyqa". The predictors of interest were Speech Style (plain, Lombard), Speaker Nativeness (native, non-native), and Listener Group (Canadian, Dutch, Spanish). The control predictors were Final, Trial Number, Occurrence, and Focus. Final (final, non-final) was included to indicate whether the target word token was produced as the final word in the sentence or not, since final words are typically lengthened and may therefore be easier to understand. Scaled and centered Trial Number was included since listeners' accuracy may improve during the experiment due to learning or drop during the experiment because of fatigue. Occurrence (first, second) indicated whether it was the listeners' first or second occurrence of hearing the target word and was included because priming could make the second occurrence easier to comprehend. Finally, Focus indicated whether the target word token came immediately after the word with contrastive focus in the original sentence in the corpus or at some later point in the sentence. Focus was included since we were unsure whether the effect of post-focus compression on intelligibility varies based on the closeness to the word with contrastive focus. Speaker, Listener, and Target Word were the crossed-random effects.

The same statistical procedure was followed as when analyzing HEGPs in determining the best model, using a backwards fitting procedure for the fixed effects structure and then a forward fitting procedure for the random slopes. We began with a theory-based approach with simple effects and interactions among our predictors of interest (Speech Style, Speaker Nativeness, and Listener Group) and simple effects for our control predictors (Final, Trial Number, Occurrence, Focus). In the model reported below, plain speech (Speech Style), non-native speakers (Speaker Nativeness), and Canadian listeners (Listener Group) are on the

intercept.

Because the Canadian listeners are on the intercept, the model does not provide detailed information about the potential differences between the Dutch and Spanish listeners. We therefore releveled the final model with the Dutch listeners on the intercept and also report the relevant results of this releveled model.

5.4.3 Results

Due to technical issues, 193 trials for the Spanish participants were lost (one Spanish participant lost 52 trials while the rest randomly lost between 0 and 9 trials; the total loss is approximately 2% of the Spanish data). This resulted in 8,832 and 8,829 data points for the Dutch and Spanish listener cohorts, respectively.

These data are visualized in Figure 2 and the final statistical model is shown in Table 2. This final model lacks some of the predictors and interactions that we tested because they were not statistically significant. While Figure 2 may suggest there is a three-way interaction among our predictors of interest (Speech Style * Speaker Nativeness * Listener Group), this was not borne out in the statistical analysis, and therefore this interaction was removed from the model (Speech Style: Lombard * Speaker Nativeness: Native * Listener Group: Dutch β= 0.2, z = 1.5, p = 1.3 and Speech Style: Lombard * Speaker Nativeness: Native * Listener Group: Spanish β= 0.1, z = 0.6, p = 0.6). The interaction of Speech Style with Speaker Nativeness was no longer significant (β= -0.2, z = -0.9, p = 0.4) after Speech Style was added as a random slope to the Speaker random intercept and therefore also removed. The control predictor Focus was also not significant and removed from the model (β= 0.1, z = 0.1, p = 0.9).

With regard to non-native speech, which is at the intercept, we found no difference between the Canadian (at the intercept) and Dutch listeners, while the Spanish listeners performed worse compared to the Canadians. A releveled model with the Dutch listeners, instead of the Canadian listeners, on the intercept, showed that the Spanish listeners also performed worse than the Dutch listeners (Listener Group: Spanish β= -0.2, z = -2.0, $p < 0.05$).

Overall the native speakers were better understood than the non-native speakers, and this was more so for the native (Canadian) listeners (interaction of Speaker Nativeness and Listener Group) than the non-native listeners. The Dutch and Spanish listeners showed a similarly sized effect of the nativeness of the speech (in the releveled model with the Dutch listeners on the intercept, the interaction of Speaker Nativeness with the Spanish Listener Group was not significant: β= 0.1, z = 0.9, p = 0.4).

Lombard Speech was better understood than plain speech. This Lombard benefit was larger for non-native listeners (interaction of Listener Group and Speech Style), with the Dutch and the Spanish listeners showing a similarly sized effect (in line with this, the releveled model with the Dutch listeners on the intercept showed no significant interaction of Speech Style with the Spanish Listener Group: Lombard $\beta = 0.0$, $z = 0.5$, $p = 0.6$). As there is no three-way interaction of Speaker Nativeness, Listener Group, and Speech Style, the effect of Lombard speech appears similar for native and non-native speech.

Regarding the control variables, when the stimulus was produced as the last word in the sentence in the corpus, participants did better, suggesting that final lengthening improved intelligibility. As trial number increased, performance improved, indicating that learning occurred over the course of the experiment. Additionally, the second occurrence of the word was better understood than the first, suggesting that priming aided the participants.

From the model's random effect structure, we learned that the intelligibility scores varied for Listeners, Target Words, and Speakers. Additionally, Speech Style affected different Target Words differently as well as affecting Speakers differently. We also observed that the different Listener Groups were affected differently by the different Target Words.

Table 2: *The glmer model of intelligibility scores for native and non-native plain and Lombard speech by native and non-native listeners. Plain speech, non-native speaker, and native listener are on the intercept. After each predictor, we indicate which level is being contrasted with the intercept. The β indicates the size of the difference between the intercept level and the contrasted level.*

Fixed Effects	*β*	*z*	*p*
(Intercept)	-1.6	-4.9	<0.001
Speaker Nativeness: Native	1.7	4.8	<0.001
Listener Group: Dutch	0.0	-0.4	0.7
Listener Group: Spanish	-0.3	-2.2	<0.05
Speech Style: Lombard	0.9	5.3	<0.001
Trial Number	0.1	4.6	<0.001
Occurrence: Second	0.4	9.2	<0.001
Final: Non-final	-0.5	-2.2	<0.05
Speaker Nativeness: Native* Listener Group: Dutch	-0.9	-11.4	<0.001
Speaker Nativeness: Native* Listener Group: Spanish	-0.8	-10.3	<0.001
Listener Group: Dutch * Speech Style: Lombard	0.2	3.1	<0.01
Listener Group: Spanish * Speech Style: Lombard	0.3	3.5	<0.001
Random Effects			*SD*
Listener (Intercept)			0.3
Target Word (Intercept)			1.5
Speech Style: Lombard by Target Word			1.3
Listener Group: Dutch by Target Word			0.4
Listener Group: Spanish by Target Word			0.7
Speaker (Intercept)			0.7
Speech Style: Lombard by Speaker			0.4

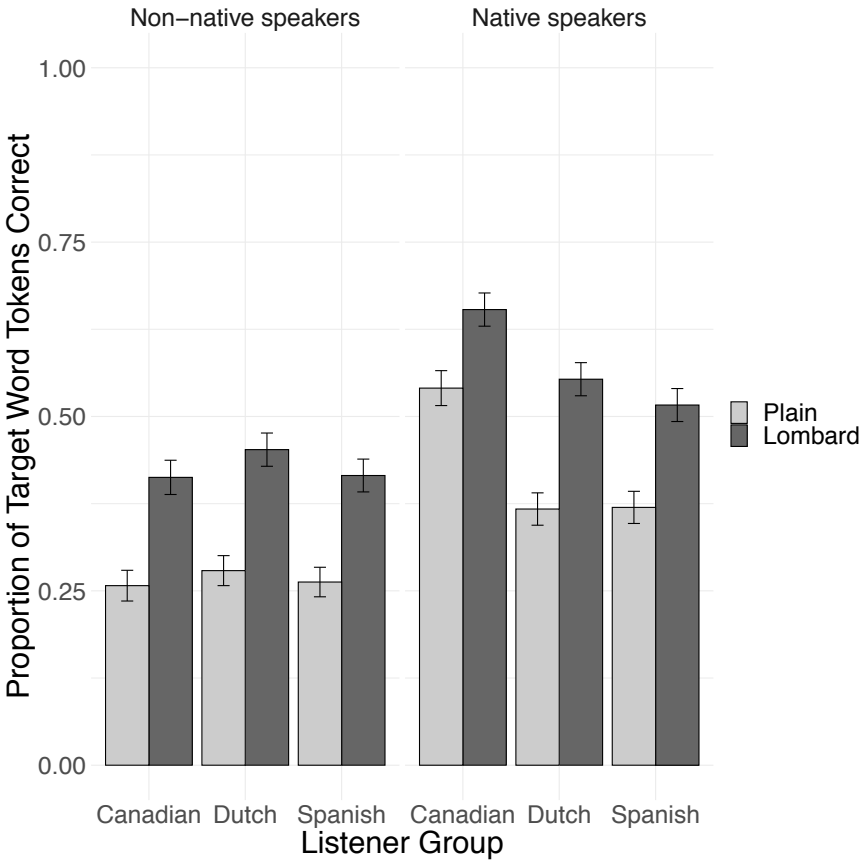

Figure 2: *The proportion of target word tokens correct for plain and Lombard speech split by Speaker Nativeness and by Listener Group. The error bars indicate 95% confidence intervals.*

5.4.4 Intelligibility analysis including HEGPs as predictor

5.4.4.1 Analyses

We extended the analysis of participants' intelligibility scores provided in section 5.4.2 by including the additional predictor HEGPs. As explained above, the HEGPs form an acoustic measure indicating the ability of a word token to resist the noise masking used in the intelligibility experiment. This additional analysis can be seen as a control analysis to see whether the effects of Speech Style and its interactions are still statistically significant after inclusion of an acoustic predictor that may partly account for the Speech Style effect.

The HEGPs from Section 5.3 were scaled and centered, because not doing so led to convergence issues. In determining the model, the same fitting procedure as in the previous analysis (Section 5.4.2) was followed. Compared to the previous analysis, we began by including the additional interaction of Listener Group with HEGPs. We did not include all interactions between the HEGPs and the variables of interest because the results from the statistical analysis of HEGPs in Section 5.3.3 showed that HEGPs are highly correlated with both Speaker Nativeness and Speech Style. We therefore excluded the interactions between HEGPs and these variables of interest and all higher order interactions containing these variables.

We established whether the model including the HEGPs explained more variance than the model without this predictor by comparing their Akaike Information Criterion (AICs). Because models can only be compared on the basis of their AIC if they are based on the same dataset, we made the comparison on the models before the removal of the residual outliers.

5.4.4.2 Results

When fitting this model, as with the previous model (Section 5.4.2), the three-way interaction (Speech Style * Speaker Nativeness * Listener Group) as well as the interaction of Speech Style with Speaker Nativeness were not significant and therefore removed from the model (Speech Style: Lombard * Speaker Nativeness: Native * Listener Group: Dutch β= 0.3, z = 1.8, p = 0.1, Speech Style: Lombard * Speaker Nativeness: Native * Listener Group: Spanish β= 0.2, z = 1.0, p = 0.3, and Speech Style: Lombard * Speaker Nativeness: Native β= -0.1, z = -0.8, p = 0.4). Additionally, the two-way interaction between Listener Group and HEGPs was not significant and therefore removed from the model (Listener Group: Dutch * HEGPs β= 0.0, z = -0.2, p = 0.9 and Listener Group: Spanish * HEGPs β = 0.0, z = -0.9, p = 0.4). The control predictor Focus was again also not significant and removed from the model (β= 0.1, z = 0.3, p = 0.7).

The final statistical model is shown in Table 3. The AIC score of the model with HEGPs (25683.6, df = 27) was substantially lower than that of the model without the HEGPs as predictor (25925.0, df = 26), which shows that the model with HEGPs better explains the data despite the extra degree of freedom.

The HEGPs contribute to explaining the variance in the data. The higher the HEGPs, the better intelligibility (main effect of HEGPs). The effect size of HEGPs is similar to the effect size of Speech Style.

The inclusion of the HEGPs changed the results for the non-native speech. In contrast to the previous model, the model with the HEGPs does not show a difference between either the Dutch or the Spanish listeners with the Canadian listeners in comprehending non-native plain

speech. The Dutch and the Spanish listeners also did not differ from each other, as evidenced by a model with the Dutch listeners on the intercept (Listener Group: Spanish β= -0.2, z = -1.9, p = 0.1).

The HEGPs model also showed different results for the native speech. Native speech was only better understood than non-native speech (simple effect of Speaker Nativeness) by the Canadian listeners. The interactions of Speaker Nativeness and both Listener Groups showed that this was less for the Dutch and Spanish listeners, and when we releveled the model, the simple effect of Speaker Nativeness was not significant (Speaker Nativeness: Native β= -0.6, z = -1.8, p = 0.1) and we did not find a difference of the effect of the nativeness of the speech between the Dutch and Spanish listeners (Speaker Nativeness: Native * Listener Group: Spanish β= 0.1, z = 0.8, p = 0.4). This indicates that if the listener is Dutch or Spanish, the benefit of the native speech can be explained by the difference in HEGPs, while the Canadian listeners benefit from the native speech even when HEGP differences are accounted for.

The inclusion of the HEGPs did not affect the pattern of results for the difference between plain and Lombard speech. Lombard speech was better understood than plain speech (simple effect of Speech Style), and this effect was larger for the Dutch and Spanish listeners (interaction of Listener Group and Speech Style) compared to Canadian listeners. The Dutch and the Spanish benefitted similarly from the Lombard speech, as indicated by the releveled model, in which the interaction of Speech Style and Spanish Listener Group was not significant (Listener Group: Spanish * Speech Style: Lombard β= 0.0, z = 0.3, p = 0.8).

The control predictors remained the same as in the model in Section 5.4.3 (Table 2), with significant effects of Final, Trial Number, and Occurrence. Furthermore, the random structure also remained the same as in the previous model.

Table 3: *The glmer model of intelligibility of native and non-native plain and Lombard speech by native and non-native listeners including HEGPs as predictor. Plain speech, non-native speaker, and native listener are on the intercept. After each predictor, we indicate which level is being contrasted with the intercept. The β indicates the size of the difference between the intercept level and the contrasted level.*

Fixed Effects	*β*	*z*	*p*
(Intercept)	-1.2	-4.2	<0.001
Speaker Nativeness: Native	1.5	4.8	<0.001
Listener Group: Dutch	0.0	-0.2	0.8
Listener Group: Spanish	-0.2	-1.9	0.1
Speech Style: Lombard	0.6	3.9	<0.001
HEGPs	0.5	15.7	<0.001
Trial Number	0.1	4.8	<0.001
Occurrence: Second	0.4	9.1	<0.001
Final: Non-final	-0.6	-2.7	<0.01
Speaker Nativeness: Native* Listener Group: Dutch	-0.9	-11.6	<0.001
Speaker Nativeness: Native* Listener Group: Spanish	-0.9	-10.6	<0.001
Listener Group: Dutch * Speech Style: Lombard	0.2	2.8	<0.01
Listener Group: Spanish * Speech Style: Lombard	0.2	3.0	<0.01
Random Effects			*SD*
Listener (Intercept)			0.3
Target Word (Intercept)			1.4
Speech Style: Lombard by Target Word			1.2
Listener Group: Dutch by Target Word			0.4
Listener Group: Spanish by Target Word			0.7
Speaker (Intercept)			0.6
Speech Style: Lombard by Speaker			0.3

5.5 Discussion

This article compares the size of the Lombard intelligibility benefit for words in noise between native (American-English) and non-native (native Dutch) English speakers, when heard by native and non-native listeners. Perception of non-native Lombard speech has not previously been investigated. This study sheds light on the nature of non-native Lombard speech while also touching upon non-native speech perception.

By computing high-energy glimpsing proportions (HEGPs), we first gained information about the speech signal itself and the stimuli's capacity to withstand noise masking. HEGPs only consider the acoustics of the speech signal and are therefore language independent and can serve as an objective measure. Previous studies have shown that a large part of the capacity to withstand masking can be explained by the shifts in the spectral energy distribution of speech (e.g., Lu & Cooke, 2009b). HEGPs only consider the quantity of audible information and not the quality. Therefore they do not take other factors that may be relevant for speech intelligibility into account, such as coarticulation and changes in vowel formants which may occur with reduction.

The native and the non-native Lombard speech showed similar increases in HEGP scores when going from plain to Lombard speech. This suggests that both native and non-native Lombard speech are more resistant to noise than plain speech. The finding for native speech is in line with previous research using glimpsing proportions (GPs), which is the basis of the HEGPs that we used (e.g., Lu & Cooke, 2009b). Importantly, the current study extends this finding to Lombard speech produced by non-native speakers. This suggests that beneficial alterations to the spectral energy distribution are also present in non-natively produced Lombard speech.

The intelligibility of the same speech materials was tested with human listeners. In addition to a native listener group (Canadians), we tested two non-native listener groups; one that shared the native language with the non-native English speakers (native Dutch) and one that did not (native Spanish). All listener groups showed a clear Lombard benefit for both the native and the non-native speech. Together, the HEGP analysis and the intelligibility experiment therefore show that, like native speakers, non-native speakers produce Lombard speech and that their Lombard speech is more intelligible in noise than their plain speech.

While our finding that native Lombard speech is more intelligible than plain speech when present in noise is in line with previous research investigating the Lombard benefit with native speakers (e.g., Dreher & O'Neill, 1957; Pittman & Wiley, 2001; Van Summers et al., 1988), to our knowledge, the Lombard intelligibility benefit with non-native speech is a novel finding. It could indicate that the production of Lombard speech is easily acquired by language learners or that it results from mechanisms that learners transfer from their native language to non-native languages. As learners are typically not explicitly taught about Lombard speech and previous research showed small differences in the median F0 and F0 range between the native and non-native English Lombard speech presented in our experiment (Chapters 3, 4), we believe that the latter explanation is more likely. This indicates that learners may implement their native Lombard speech alternations in non-native languages, but that this does not greatly hinder the Lombard benefit for the listeners.

The native listeners exhibited a smaller Lombard benefit than the non-native listeners. The increased intelligibility of Lombard speech compared to plain speech across native and non-native speech was 13.4% points for the native listeners, while, for the non-native listeners, this was greater at 16.5% points. It should be noted that although the non-native listeners may have had more problems than the natives with correctly spelling the target words (we counted every misspelling as an incorrect answer), this should not have modulated the size of the Lombard benefit, as we may expect the same spelling problems for both the Lombard and the plain speech.

Our finding that non-native listeners showed a larger Lombard benefit than native listeners contrasts with the findings reported by Cooke and García Lecumberri (2012), who documented a smaller Lombard benefit for non-native listeners compared to the same materials with native listeners (Lu & Cooke, 2008). Lu and Cooke (2008) reported native listeners identifying plain speech embedded in noise at 42% accuracy, with a 22% point increase in intelligibility for identifying Lombard speech produced in 82 dB SPL embedded in noise. For non-native listeners, this was a baseline of 36% accuracy and a 15% point increase in intelligibility for Lombard speech (Cooke & García Lecumberri, 2012). This difference with our findings could be explained in several ways, including differences in stimuli and masker properties. The listeners tested in Cooke and García Lecumberri (2012) and in Lu and Cooke (2008) identified letter and number combinations at SNR -9 dB. In the current study, listeners were asked to identify English words at -1 SNR. For our stimuli, we used 96 different target words, which did not belong to any one category or topic. Our stimuli were produced in sentences where many words, including the target words, were reduced (e.g. *police* was often pronounced as /pli:s/). We spliced the target words out of their sentences and presented them in isolation, which makes especially reduced words hard to understand (e.g., Ernestus, Baayen, & Schreuder, 2002). The difference in stimuli and masker between our study and Cooke and García Lecumberri (2012) and Lu and Cooke's (2008) could in part explain the difference in the size of the Lombard benefit.

We also analyzed the intelligibility scores including HEGPs as a predictor. The predictor HEGP was statistically significant, suggesting that part of the Lombard benefit can be explained by a shift in energy to higher frequency regions. Since Speech Style was still significant as well, these analyses show that the HEGPs predict part of the Lombard benefit while other acoustic characteristics of Lombard speech not captured by HEGPs also contribute to the Lombard benefit. Acoustic characteristics of Lombard speech that are not considered in HEGPs include increase in duration (e.g., Castellanos et al., 1996; Dreher & O'Neill, 1957; Garnier & Henrich, 2014; Junqua, 1993; Van Summers et al., 1988) and shifts in the vowel space (e.g., Garnier, 2008). The inclusion of the HEGPs as predictor did not affect any of the interactions involving Speech Style.

In addition to observing the Lombard benefit in both native and non-native speech for all listener groups, we obtained results further elucidating the differences between native and non-native speech and between native and non-native listening. The Spanish listeners performed worse than the Canadian native and Dutch non-native listeners when listening to non-native plain speech. Dutch listeners may outperform Spanish listeners because of their greater exposure to English in daily life. Perhaps more interestingly, when we compared the model without the predictor HEGPs with a model with the predictor HEGPs, the latter model showed a better fit with the data and showed that the Spanish performed as well as the Dutch and native listeners for non-native plain speech. This difference between the two models in whether the Spanish listeners performed as well as the other two groups suggests that the Spanish listeners are differently affected than the native and Dutch listeners by the energetic masking as indicated by the HEGPs. As this experiment was not designed to test for possible differences among groups in sensitivity to HEGPS, future research should further investigate this possible difference.

All listeners benefitted from listening to native rather than non-native speech. This is not in line with the matched and mismatched interlanguage speech intelligibility benefit, as the Dutch and Spanish listeners did not find the non-native English (native Dutch) speech as intelligible as the native speech (Bent and Bradlow, 2003). The higher intelligibility for native speech mimics the difference in HEGPs between the native and the non-native speech indicating that native English speech better withstood masking. Unfortunately, we cannot establish whether this difference in HEGPs is due to acoustic properties inherent to native versus non-native speech or whether it is due to differences between English and Dutch (such as the different realizations of fricatives) with the Dutch characteristics surfacing in non-native English produced by the native Dutch speakers. This finding requires further investigation.

The Canadian native listeners benefited more from listening to native speech compared to the Spanish and the Dutch. Stated differently, the Canadian listeners suffered more from listening to non-native speech than the Dutch and the Spanish listeners. When the HEGP metric was incorporated in the analysis, the difference between the native listeners on the one hand and the non-native listeners on the other hand was larger, with the Dutch and Spanish listeners no longer showing a benefit for the native speech. Therefore, when the HEGPs are taken into consideration, the Dutch and Spanish listeners do show a matched and mismatched interlanguage speech intelligibility benefit respectively (Bent & Bradlow, 2003). This suggests again that different listener groups may benefit differently from the energetic masking as indicated by the HEGPs.

The materials in this study were restricted to native English and non-native English produced by native speakers of Dutch and we only tested native listeners of English, Dutch, and Spanish. The choice of these languages and these listeners may have affected the results:

different results may have been obtained had we chosen to study speakers of native languages that are more dissimilar than English and Dutch, and listeners of native languages that are more dissimilar than English, Dutch, and Spanish. Considering there are various factors that influence the matched and mismatched interlanguage speech intelligibility benefit (Bent & Bradlow, 2003), including the language choices of the native and non-native speakers, the proficiency of the non-native speakers (e.g., Stibbard & Lee, 2006), and the proficiency of the non-native listeners (Imai, Walley, & Flege, 2005; Pinet, Iverson, & Huckvale, 2011), we leave it to future research to investigate to what extent the results obtained in our study generalize to other languages and listener groups. Especially, the acoustic characteristics not captured by HEGPs may differ more among languages that are less similar to each other than Dutch and English are. If so, the Lombard benefit may depend on the combination of the speaker's and the listener's native languages.

In this study, we set out to examine the size of the Lombard intelligibility benefit for native and non-native speech. We approached this by analyzing HEGPs to understand the speech signal itself and by conducting an intelligibility experiment with native and non-native listener groups to also understand the role of the listener's native language. We found that, like native speakers, non-native speakers can produce Lombard speech with higher HEGPs than plain speech and that is clearly beneficial for the listener. Although there may be influences from the speaker's native language on non-native Lombard speech (Chapters 3, 4), non-native speech can still show a Lombard intelligibility benefit.

Chapter 6:

How articulatory effort affects the strength of non-native speakers' accentedness

Abstract

We investigated how articulatory effort affects within-speaker variation in the degree of the non-native accent. Native Dutch and English speakers produced English target words in two distinct conditions: in noise with contrastive focus on the word (greater effort: GE) and in quiet with no contrastive focus (less effort: LE). We focused on phonemes in the target words that the Dutch speakers of English have difficulties with (/θ/, schwas that correspond to unstressed full vowels in Dutch, and final voiced obstruents). A forced aligner gauged whether the phone of interest was pronounced in a prototypical English manner, finding that the native and non-native speakers paralleled each other, only showing a difference between GE and LE for the schwa category, with more schwas in the LE condition. Additionally, native and non-native listeners listened to the two realizations of the words by the same speaker and indicated which was more native-like. Native listeners tended to prefer the LE realization for native speakers, but they did less so for non-native speakers. The non-native listeners did not show preferences for either speaker group. While for most words it does not matter whether non-native speakers produce them with much or little effort, for some words much articulatory effort makes them sound less native like.

This chapter is an edited version of:

Marcoux, K., Süß, E., & Ernestus, M. (to be submitted). How articulatory effort affects the strength of non-native speakers' accentedness.

6.1 Introduction

One of the most difficult features to master in a non-native language is its pronunciation. Speakers who are fluent in the non-native language's syntax, morphology, and pragmatics are often still identifiable as non-native speakers due to their non-native accent (Flege, 1995). The non-native speakers may show the influence of their native languages in, for instance, their pronunciation of the segments that do not occur in their native languages' phoneme inventories, their prosody (e.g., deviant lexical stress patterns), or the application of phonological processes (e.g., final devoicing). Native listeners easily recognize the smallest deviations from native speech, in as little as 30 milliseconds (Flege, 1984). While many studies have investigated predictors of a learner's success in acquiring a native-like pronunciation, thus explaining variation *among* speakers, very few studies have investigated *within-speaker* variation in the degree of the non-native accent. Such studies will provide us with more insight into the mechanisms responsible for a non-native accent.

According to Piske, MacKay, and Flege (2001), the numerous studies investigating the factors that contribute to differences *among* speakers in the degree of their non-native accent show that the most important predictor is the age at which the second language (L2) is acquired: the earlier this happens, the less severe the ultimate degree of non-native accent (e.g., Asher & García, 1969; Flege, Munro, & Mackay, 1995; Oyama, 1976; Piske et al., 2001; Thompson, 1991). Additional predictors include length of residence where the L2 is spoken (e.g., Asher & García, 1969; Flege et al., 1995), the type of phonetic training in the language classroom (e.g., Missaglia, 1999), gender (although its effects are not univocal; e.g. Asher & García, 1969; Flege et al., 1995; Thompson, 1991), motivation (e.g., Purcell & Suter, 1980), the use of the native language (L1) and L2 (e.g., Flege et al., 1995; Purcell & Suter, 1980), and the ability to mimic, a measure of language learning aptitude (e.g., Purcell & Suter, 1980; Thompson, 1991).

One of the few studies addressing which factors contribute to *within-speaker* variation, compared the degree of a speaker's non-native accent in different tasks (Gustafson, Engstler, & Goldrick, 2013). The authors compared a repetition task with a picture naming task, which involves semantic processing, finding that the non-natives were slower and more accented when performing the picture naming task compared to the repetition task. From this, they discuss the possibility that when the task involves semantic processing, cognitive resources are used and therefore these resources may be less available for phonetic processing. Other research compared read speech to spontaneous speech, and found that read speech was judged as having a stronger non-native accent (e.g., Oyama, 1976; Thompson, 1991). However, as Piske and colleagues (2001) caution, in reviewing these articles, each of the tasks pose different confounds; reading abilities will affect the read speech, while disfluencies will

affect the spontaneous speech. Additionally, with spontaneous speech, speakers can avoid L2 phones that are difficult for them. It is therefore not yet clear which factors lead to more non-native like pronunciations.

In the present study, we further investigate what may cause *within-speaker* variation in the degree of the non-native accent by focusing on the role of articulatory effort. If non-native speakers know how to produce the non-native sounds, we may expect that more articulatory effort may lead to a weaker native accent. In contrast, if non-native speakers may not entirely know how to articulate the non-native sounds, we may expect the reverse, and that more articulatory effort may lead to a stronger non-native accent. This latter hypothesis is in line with the finding reported above that read speech may lead to a stronger non-native accent than spontaneous speech (e.g., Oyama, 1976; Thompson, 1991), where speakers may focus more on content than on a proper pronunciation. Which of the two hypotheses is correct may differ per language feature (e.g., among phonemes) depending on the phonological similarities between the native and non-native language.

If speakers produce a non-native accent to a larger degree when increasing articulatory effort for pronunciation, then greater effort is counterproductive. The degree of the accent has social implications as judgements are often made about the speaker based on their accent, whether regional (e.g., Mulac & Rudd, 1977) or non-native (e.g., Munro, Derwing, & Sato, 2006). Native listeners often judge non-native individuals' personality and intelligence more negatively (e.g., Fuse, Navichkova, & Alloggio, 2018; Tsurutani, 2012), as less credible (e.g., Lev-Ari & Keysar, 2010), and as a better fit for low-status than high-status jobs (e.g., Kalin & Rayko, 1978).

These negative judgments of non-native speakers are not limited to native listeners. Non-native listeners also judge other non-native speakers negatively. Like native listeners (Lev-Ari & Keysar, 2010), non-natives listeners tend to find non-native speech less credible (Hanzlíková & Skarnitzl, 2017). When it comes to educational settings, non-native students evaluate moderately accented non-native lecturers as less competent than lecturers with a slight non-native or with a native accent (e.g., Hendriks, van Meurs, & Hogervorst, 2016; Hendriks, van Meurs, & Reimer, 2018). Similarly, non-native evaluators find job-candidates with a strong non-native accent as less hirable compared to the slight non-native or the native accent (Roessel, Schoel, Zimmermann, & Stahlberg, 2019).

In this article, we investigate how the degree of the non-native accent is affected by the articulatory effort speakers invest in their pronunciation by comparing target words produced in noise and with contrastive focus (Greater Effort condition; GE), where we expect increased effort, with words produced in quiet without contrastive focus, where we expect less effort (Less Effort condition; LE). Several studies have shown that speakers invest more vocal and

articulatory effort in words with contrastive focus than those without contrastive focus. Choi (2003), for instance, showed that the contrast between /p/ and /b/, as produced by native English speakers, is larger for words that carry contrastive focus in the sentence than for words that do not, as indicated by voice-onset-time (VOT) measurements for the six participants and by the fundamental frequency (F0) measurements for four of the six participants. Choi's findings from the laboratory were supported in an analysis of news speech read over the radio (Cole, Kim, Choi, & Hasegawa-Johnson, 2007). The authors found a general contrast enhancement for voicing cues for English stops under focus: phrasal accent increased VOT, F0, and closure duration, affecting voiced and voiceless stops to a different extent. Together, these articles demonstrate that, when a word is under focus, generally features are enhanced, making the phones more distinct from similar phones.

Vocal effort is increased for speech produced in noise, Lombard speech. Lombard speech has a higher fundamental frequency (F0) and a higher first formant (F1) as well as more energy in higher frequencies compared to plain speech (e.g., Junqua, 1993; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). According to Schulman (1989), the higher F0 and F1 are characteristic of increased vocal effort. Further, in Lombard speech vowels and consonants are lengthened (but vowels more than consonants; e.g., Castellanos, Benedí, & Casacuberta, 1996; Fónagy & Fónagy, 1966; Garnier & Henrich, 2014; Junqua, 1993). Some of the acoustic modifications of Lombard speech affect linguistic information. Examples are, increase in the first formant (F1), which may lead to shifts in vowels (e.g., Garnier, 2008; Godoy, Koutsogiannaki, & Stylianou, 2014; Mixdorff, Pech, Davis, & Kim, 2007), changes in F2 that may enhance or rather reduce vowel contrasts (Perkell et al., 2007), and smaller within-category vowel variation (Cooke & Lu, 2010).

In our study, native (American-English) and non-native English speakers, who were native speakers of Dutch, read English sentences with target words embedded. We chose target words that are challenging for this group of non-native English (native Dutch) speakers, increasing the chance of seeing pronunciation variation in the two articulatory effort conditions. Our three target word categories were: /θ/-initial words, Dutch-English cognates with schwa in pre-stress position in English and a full vowel in its place in the Dutch cognate, and words with final voiced obstruents (hereafter referred to as /θ/, schwa, and voiced obstruent target word category, respectively).

Many Dutch speakers have difficulties with /θ/-initial words as /θ/ does not exist in their native phonemic inventory. Accordingly, Gussenhoven and Broeders (1997) address /θ/ as a problematic phoneme in their guidebook "English pronunciation for student teachers". When Dutch students learn English in school, they are explicitly made aware of the pronunciation difficulties of /θ/. Nevertheless, Hanulikova and Weber (2010) found that Dutch university students only pronounce [θ] correctly 62% of the time, and that the most frequent substitution

is [t], accounting for 77% of the substitutions. However, they also found that most Dutch speakers are not limited to one substitution type (e.g., [s] and [f]).

We hypothesize that, for the /θ/ target words, an increase in effort (the GE condition) will result in a more native like pronunciation. When speakers are aware that they have difficulties pronouncing a certain sound, such as is the case with /θ/, their increased effort may result in a more distinct pronunciation of that sound, which may be closer to the native pronunciation. We therefore expect that the Dutch speakers will produce more instances of /θ/ in the GE condition than in the LE condition.

The schwa target words are Dutch-English cognates where the English pronunciation has schwa in pre-stress position while in the Dutch pronunciation there is a full vowel in its place (e.g., the English word *banana* /bə'nana/ versus Dutch *banaan* /ba'nan/). In English, unstressed vowels are typically reduced to schwa, while this reduction is optional in Dutch (e.g., Kager, 1989). In Dutch, the reduction of a full vowel to schwa is not only influenced by lexical stress but also by whether the word has contrastive focus, with words with contrastive focus eliciting fewer schwa productions (e.g., van Bergem, 1993).

We hypothesize that when Dutch speakers spend more effort on their English pronunciation, they are more likely to incorrectly produce the unstressed vowels as full vowels, as they do in their native language. In contrast, less articulatory effort (LE) for Dutch speakers will result in more schwa realizations. That is, we expect more native like pronunciations in the LE than in the GE conditions. In defining this category of target words, we focused on target words where the vowel was orthographically written with <a> or <o> (e.g., *banana* and *police*) and not as an <e> as that would encourage the realization of schwas.

Final voiced obstruents are problematic for Dutch speakers because of final devoicing in Dutch, which implies that the final voiced obstruent is devoiced (e.g., Berendsen, 1986; Booij, 1985; Simon, 2010). English, on the other hand, does not have final devoicing as shown by minimal word pairs such as *bad-bat* (/bæd/ - /bæt/). This means that the final voiced obstruents in our English target words may be devoiced by the Dutch speakers. For this target word category, we selected words ending in the voiced obstruents /b/ and /d/.

We do not have a clear hypothesis for the voiced obstruent target words. We have the impression that Dutch speakers are not generally aware that final devoicing is a feature of their language and hence are probably not aware that they incorrectly devoice final voiced obstruents in English. This means that we do not expect a change in their non-native pronunciation when they increase their effort.

In order to gauge what our native and non-native speakers produced, we first had a forced speech aligner, Montreal Forced Aligner (MFA; McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017) produce phonetic transcriptions of the words. We used a forced aligner

as it is objective, consistent and not susceptible to fatigue, as human transcribers are. We examined MFA's phone level transcription to see whether, how and when the speakers mispronounced /θ/, schwa, and the voiced obstruents. These analyses provide us with objective measures indicating the presence or lack thereof of the native like phoneme, which in turn signals a native-like accent or not.

We additionally conducted a perception experiment where native and non-native listeners heard the same stimuli as transcribed by the MFA, presented in quiet. In each trial, participants heard the two realizations of a target word, both produced by the same speaker, one from each articulatory effort condition (GE and LE) consecutively. The participants' task was to indicate which realization sounded more native-like. While the phonetic transcriptions provided us with information about whether the phonemes we are interested in were produced in a prototypical American-English pronunciation (native-like), the perception results showed how native like the *complete* word sounds to human listeners. We use the phonetic transcriptions when interpreting listeners' evaluations of the nativeness of the LE and GE conditions in determining whether the difference may result from a shift in phoneme category. Across all three target word categories, we expect to find a general preference for the plain speech (LE), since Lombard speech (GE) may sound unnatural when heard in quiet, as its acoustic modifications are made in response to noise. For our research question, we were interested in how this difference between Lombard and plain speech in perceived accentedness varied per target word category.

Since it is feasible that characteristics of a listeners' native language may influence the accentedness interpretation, having native and non-native listeners would allow for a comparison between the two groups. We presented native and non-native speech to native (Canadian) and non-native (native Spanish or Basque/Spanish bilingual listeners furthermore referred to as our Spanish listeners) listeners. In terms of the target word categories chosen, like English, peninsular Spanish has the phoneme /θ/ (e.g., Hualde, Olarrea, Escobar, & Travis, 2001). The two listener groups may therefore be similar in terms of their evaluation of the pronunciation of the /θ/. In contrast, Spanish differs from English in that it does not have schwa in its phonemic inventory (e.g., Lacabex, García Lecumberri, & Cooke, 2005).

6.2 Experiment 1: Computer Evaluation

6.2.1 Speech Materials

6.2.1.1 Speakers

The target words came from Chapter 2's Dutch English Lombard Native Non-Native (DELNN) corpus, for which nine native American-English and thirty native Dutch female speakers were recorded while they read sentences in quiet and in noise. From these 39

speakers in the corpus, we took a subset of 16 speakers: eight native American-English and eight native Dutch speakers. We excluded the American-English speaker who did not give permission to have her recording used in online experiments.

We selected the eight Dutch speakers from the 23 who agreed to have their recordings used online, on the basis of an overall accentedness rating experiment. Six native American-English listeners rated 150 randomly selected sentences produced in quiet (six sentences per speaker, with two sentences from each of the target word categories for each speaker) from the 23 Dutch speakers and two American-English speakers. We included two American-English speakers to confirm that the listeners were able to detect native speech. The listeners rated the overall accentedness of each sentence on a 7-point Likert scale (1 "native-like" to 7 "very strong foreign accent"). The sentences were blocked by speaker, so the listeners could adjust to each speaker. We averaged the ratings across all listeners for each speaker, and observed a median rating of 4.7 (ranging from 2.4 to 5.9) for the non-native speakers. The two American-English speakers had ratings of 1.00 and 1.03, confirming that they were correctly identified as native speakers. We chose the eight mid-accented Dutch speakers with accentedness ratings ranging from 4.4 to 5.0, because we assumed that these speakers may show the largest difference in the pronunciation of the three types of target words in the GE and LE conditions.

6.2.1.2 Stimuli

The speakers in the corpus recorded sentences in a sound attenuated room using a Sennheiser ME 64 or 65 microphone, which fed into an AudiTon preamplifier and then to a Roland R-05 WAVE/MP3 Recorder. The resulting wav file had a 44.1 kHz sampling rate with 16-bit resolution. Speakers wore a pair of Sennheiser HD 215 MKII DJ headphones, through which nothing was played when eliciting plain speech, while they heard Speech Shaped Noise at 83dB SPL through them when eliciting Lombard speech. The noise level was calibrated using the Brüel & Kjær Type 4153 artificial ear.

For the current study, we used 28 of the original 36 target words in the three categories. We eliminated the five target words that, when pronounced in a Dutch accent, may be homophonous with real English words (*pub, cab, lab, food,* and *theme)*. For instance, the word *pub*, pronounced /pʌb/ in American-English, may be pronounced as *pup* /pʌp/ in a Dutch accent because of final devoicing and *theme* /θim/ may be pronounced as *team* /tim/ in a Dutch accent. We also removed three target words that often had dysfluencies (*thermometer*, *thermodynamics*, and *massage*). See Appendix 1 for the list of target words used as well as the Dutch translations. This table illustrates that all /θ/ and schwa words are cognates, and that half of the voiced obstruent words are cognates as well.

The speakers produced the target words in answers to questions in two extreme effort conditions. In one extreme, they heard noise and placed contrastive focus on the target word (GE). In the other extreme, they produced the sentence in quiet and the contrastive focus was placed elsewhere in the sentence (LE). An example of a contrastive question-answer pair for the GE condition is found below in which the target word *parade* has contrastive focus, as indicated by the underlining.

1. *Did the family go to the beach in Barcelona?*
   No, they went to the parade in Barcelona.

Next is an example of the LE condition, where the target word, *parade*, does not receive contrastive focus, but rather the contrastive focus comes earlier in the sentence.

2. *Did the friends go to the parade in Barcelona?*
   No, the family went to the parade in Barcelona.

The target words were taken from the answers of these question-answer pairs. We had the 28 target words produced in both articulatory effort conditions by each of the 16 speakers, resulting in 896 tokens. Ten tokens were excluded because of dysfluencies of five tokens by the participant in one condition, for which we also excluded the corresponding target word in the other condition (*thriller*, *theology*, *theta*, *botanical*, and *thermal* were removed for one native Dutch speaker each). This resulted in a total of 886 tokens.

6.2.1.3 Methods

The audio was segmented at the sentence level, cutting each question and answer separately. In order to obtain word- and phone-level transcriptions aligned with the speech signal as accurately and efficiently as possible, we incorporated any dysfluencies and false starts into the orthographic transcription of the answers. We used the Montreal Forced Aligner (MFA) for the forced speech alignment, which uses Kaldi as its basis (McAuliffe et al., 2017). We provided MFA with the speech signal and the corresponding orthographic transcription. MFA takes the words in the orthographic transcription and looks for them in the pronunciation dictionary, where each word has one or more pronunciation variant. In addition, MFA needs phone models, which link the (symbolic) phone symbols in the pronunciation dictionary with the acoustic statistical properties of each phone. The MFA output is a segmentation and labelling of the input audio file at the word and phone level.

For the lexicon we used the Carnegie Mellon University (CMU) Pronouncing Dictionary, which has American-English pronunciations of words (*CMU Pronouncing Dictionary*, 2015). Since the aim of this transcription was to find out to what extent the target words were produced with a Dutch accent, we added separate Dutch-accented pronunciations of the 28 target words to the CMU Pronouncing Dictionary. This allowed the forced aligner to choose between the American-English and Dutch pronunciations of the target word. For the /θ/ target words, we allowed for the alternative pronunciations of /θ/ as /t/, /d/, /f/, /v/, /s/, and /z/. For the schwa target words, we allowed variation based on the orthographic spelling of the word. When the schwa was spelled with an <a>, we additionally allowed it to be transcribed as /ʌ/, /æ/, /ɑ/, and /ɔ/, whereas for <o> we allowed /ɔ/, /o/, and /ɑ/. For the voiced obstruent target words we allowed /d/ and /b/ at the end of the target words to be pronounced as /t/ and /p/, respectively.

We used the English acoustic models from the MFA website (McAuliffe et al., 2017), which were trained on the LibriSpeech corpus, consisting of 1,000 hours of read speech (Panayotov, Chen, & Povey, 2015). MFA first uses monophone Gaussian Mixture Models-based (GMM) Hidden Markov Models (HMMs) and then it trains triphone GMM-HMM models, which take into account the context of the phone. The acoustic models used included 68 GMM-HMMs for vowel and consonant phonemes as well as GMM-HMMs for silence. MFA uses 13 mel-frequency cepstral coefficients (MFCCs) as well as 13 more MFCCs for delta and delta-delta each, for a total of 39 features per frame. After MFCC calculation, MFA applies cepstral mean and variance normalization (CMVN) per speaker, to account for channel and recording conditions. For the CMVN to work best, we specified the number of speakers. MFA allows for speaker adaptation, which we disabled because we wanted to be able to directly compare the Dutch-accented English alignment with the American-English alignment using the same acoustic models.

In order to evaluate MFA's accuracy in transcribing the phones of interest, we had two native English listeners annotate 50 phones of each category of interest (/θ/, schwa, voiced obstruents) for a total of 150 phones. The results of the annotations can be seen in Table 1 below. In each category, the agreement was consistently higher among human transcribers, but MFA still performed sufficiently. It should be noted that identifying vowels is quite difficult which resulted in disagreement between the human transcribers, as well as between the humans and MFA. The difference found between the humans and MFA for the voiced obstruent category is likely related to the difference in weighting of cues between human listeners and the forced aligner. It is known that native American-English listeners heavily rely on the duration of the previous vowel when determining voicing at the end of a word (e.g., Raphael, 1972), while MFA relies more on cues in these obstruents themselves.

Table 1: *The percentage agreement among MFA and human annotators 1 and 2 for the phone of interest in each target word category.*

Category	Comparison	Percentage Agreement
/θ/	MFA - Human 1	84 %
	MFA - Human 2	84 %
	Human 1 - Human 2	96 %
schwa	MFA - Human 1	54 %
	MFA - Human 2	56 %
	Human 1 - Human 2	62 %
voiced obstruent	MFA - Human 1	74 %
	MFA - Human 2	76 %
	Human 1 - Human 2	90 %

6.2.2 Analysis

From the MFA alignment, we extracted the phone of interest in the target words (initial consonant in the /θ/ target words, the first vowel in the schwa target words, and final obstruent in voiced obstruent target words). We scored the phone on whether it was the prototypical American-English pronunciation (/θ/, schwa, and voiced obstruent, respectively) or whether it was pronounced as one of the Dutch accented variations we allowed.

We analyzed the MFA phone transcription using generalized linear mixed effect models (glmers) with the binomial link function and the "bobyqa" optimizer to increase the maximum number of iterations, in R version 3.5.1 (R Core Team, 2016) using the *lme4* package (version 1.1.2; Bates, Mächler, Bolker, & Walker, 2015). The dependent variable was whether the phone was the prototypical English pronunciation or not. The fixed effects were the Target Word Category (/θ/, schwa, and voiced obstruent), Speaker Nativeness (native, non-native), and Effort (GE, LE). Our control predictor was scaled and centered Trial Number to account for speakers' fatigue or practice effects. We included Target Word and Speaker for our random effects.

We used a theory-based model, beginning with interactions among our predictors of interest (Target Word Category, Speaker Nativeness, and Effort). If an interaction was not significant ($p \geq 0.05$) it was removed from the model. Simple effects were only removed if they were neither significant nor part of a significant interaction. Once the fixed effects structure was determined, we checked for the significant predictors as random slopes. We

used anovas to determine if the addition of a random slope to the random structure improved the model. For the final model, we removed absolute standardized residuals above 2.5 from the previous model, and refitted the model resulting in the final model reported below (this did not affect the patterns observed). We confirmed the significance of all predictors in the final model using the summary() function as well as the Anova() function from the *car* package (version 3.0.6; Fox & Weisberg, 2019) which performs Type II Wald chi-square tests.

In the model reported below, the schwa target words (Target Word Category), native speaker (Speaker Nativeness), and LE (Effort) are on the intercept. We had native speaker and the LE condition on the intercept since we consider these our baseline. We chose the schwa target word category as our baseline because for the human evaluation (Experiment 2), that is where we expected a difference between the Spanish and Canadian listeners. In order to best compare the computer evaluation (Experiment 1) with the human evaluation (Experiment 2), we chose to have the schwa target word category on the intercept here as well.

6.2.3 Results

MFA's phone transcription is visualized in Figure 1 and the statistical model is presented in Table 2. There is a significant interaction of Target Word Category with Effort; therefore to better interpret the model we split the data by Target Word Category, using the same fitting procedure as for the overall model.

Table 2: *Table of Type II Wald Chi-square Tests of the glmer model of MFA's labeling of the target phones. Significance values (p) are indicated using the following ranges >0.1, <0.1. <0.05, <0.01, and <0.001, where anything <0.05 or below is considered significant.*

Fixed Effects	*Chi-square*	*p*
Speaker Nativeness	66.4	<0.001
Target Word Category	1.5	>0.1
Effort	9.1	<0.01
Target Word Category * Effort	11.3	<0.01

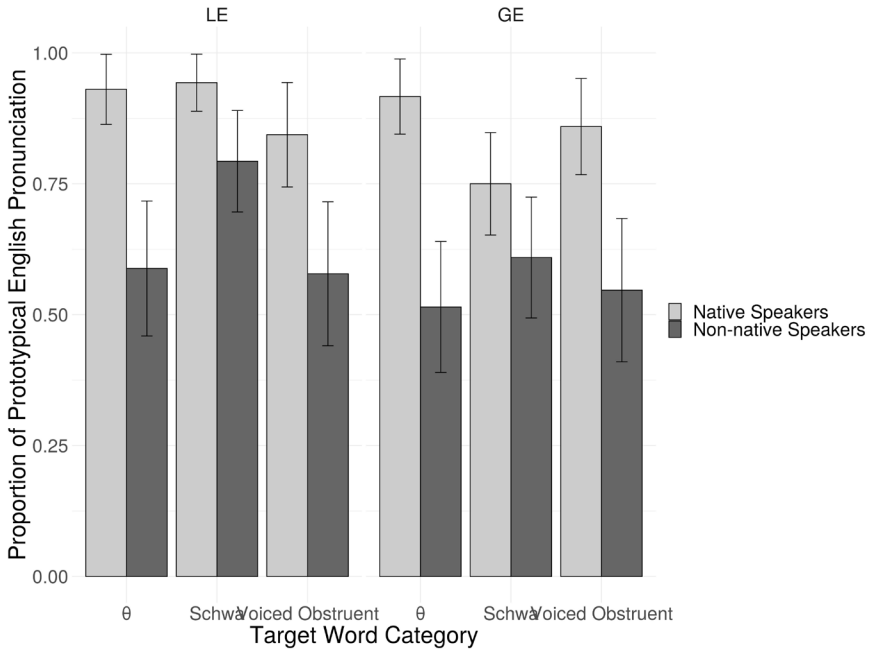

Figure 1: *MFA's labeling of the phones in the Target Word Categories split by Speaker Nativeness and by Effort Condition. The bar plots include indications of 95% confidence intervals.*

The statistical model for each target word category can be seen in Table 3. For both /θ/ and voiced obstruent target words, we only found the significant simple effect of Speaker Nativeness, meaning that the non-natives produced fewer /θ/ and fewer voiced obstruents than the native speakers. While the majority of both native and non-native speakers produced /θ/ correctly (see Table 4), native speakers produced /θ/ more often (92.4%) than non-native speakers (55.2%). Non-natives speakers' most common substitution was /t/ (28%). With the /θ/ target words, we thus did not observe the expected interaction of Speaker Nativeness with Effort. As for the voiced obstruents, MFA's transcription indicated that the non-native speakers devoiced more (43.8%) than the native speakers (14.8%). For the schwa target word category, we observed significant simple effects of Speaker Nativeness and Effort. The expected interaction of Speaker Nativeness and Effort was not found. These results mean that fewer schwas were produced when the speaker was non-native and when it was the GE condition. The native speakers produced schwas 94.3% of the time in the LE condition and 75% in the GE condition while the non-native speakers produced schwas 79.3% of the

time in the LE condition and 60.9% of the time in the GE condition. For all three Target Word Categories, MFA's labeling of phones differed per Speaker as well as per Target Word (significant random intercept of Speaker and Target Word).

Table 3: *The glmer models per target word category of MFA's labeling of target phones. Significance values (p) are indicated using the following ranges >0.1, <0.1. <0.05, <0.01, and <0.001, where anything <0.05 or below is considered significant.*

	/θ/			Schwa			Voiced obstruents		
Fixed Effects	*β*	*z*	*p*	*β*	*z*	*p*	*β*	*z*	*p*
(Intercept)	3.2	4.75	<0.001	3.9	5.15	<0.001	2.1	5.24	<0.001
Speaker Nativeness: Non-native	-2.9	-3.92	<0.001	-1.7	-3.37	<0.001	-1.8	-4.14	<0.001
Effort: GE	-	-	-	-1.6	-4.38	<0.001	-	-	-
Random Effects			*SD*			*SD*			*SD*
Speaker (Intercept)			1.16			0.62			0.52
Target Word (Intercept)			0.97			1.74			0.54

Table 4: *MFA's transcription of /θ/ target words produced by native and non-native speakers. Note that the percentages do not add to 100% because they were rounded off.*

	Non-Native English		Native English	
Phone	Count	Percentage	Count	Percentage
/θ/	75	55.2 %	133	92.4 %
/t/	38	27.9 %	0	0.0 %
/s/	12	8.8 %	0	0.0 %
/d/	7	5.2 %	3	2.1 %
/f/	3	2.2 %	8	5.6 %
/z/	1	0.7 %	0	0.0 %
Total	136		144	

In summary, both the native and the non-native speakers did not significantly adapt their pronunciation of /θ/ and voiced obstruents in GE. We only observed a difference in pronunciation between GE and LE for the schwa target words, for both speaker groups.

6.3 Experiment 2: Human Evaluation

6.3.1 Methods

6.3.1.1 Participants

Forty-four native English speakers from the University of Alberta, Edmonton, Canada (29 females, 14 males, and one other, average age (*M*) = 19.6 years, standard deviation (*sd*) = 1.8 years) completed the experiment. Participants had limited familiarity (beginner level or none) with a language that has final devoicing or that does not include schwa in its phonemic inventory.

In addition, 39 native Spanish speakers (31 females, seven males, and one other, *M* = 20 years, *sd* = 2.6 years) from the Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), in Vitoria, Spain participated in this experiment. At the time of participation, participants were completing their studies in "English Studies" and had an average LexTALE score (Lemhöfer & Broersma, 2012) of 69.94, which corresponds to an English level of B2 in the Common European Framework. All participants (Canadian and Spanish) gave informed consent and were given course credit or paid in exchange for participating in the experiment.

6.3.1.2 Speech Materials

The same speech materials (speakers, target words, articulatory effort conditions) as were used in the computer evaluation (Experiment 1) were used for the human evaluation. MFA provided the phone and word level transcriptions of the sentences produced (McAuliffe et al., 2017). The authors checked the alignment of the target words in the carrier sentences and the target word boundaries were placed at zero-crossings before being extracted and normalized at 70dB. The two realizations of the target word – the one produced in the LE condition and the one in the GE condition – by the same speaker were concatenated with one second of silence in between them, resulting in one set. We created two orders for every set (order 1: GE-LE; and order 2: LE-GE). This resulted in 896 tokens (2 orders * 16 speakers * 28 target words). The same ten tokens that were excluded for the computer evaluation (Experiment 1) were excluded here (one set of *thriller*, *theology*, *theta*, *botanical*, and *thermal* each by a non-native speaker that had issues, see Experiment 1). This resulted in 443 sets for the experiment.

6.3.1.3 Stimuli Lists

Each participant started with four practice sets not containing target words (*girl*, *sister*, *job*, *parents*) produced by a native American-English speaker who was not included in the DELNN corpus (Chapter 3). The experiment itself consisted of the 443 sets, half presented in order 1 and the other half in order 2. The stimuli were blocked by speaker, so the listeners could adjust to each speaker and the order of the target words in a block was pseudo-randomized such that no more than three target words from the same category followed each other. We had 12 lists, which randomized the order of the speakers, ensuing that no more than three native or three non-native speakers were in succession. Every participant was presented with one list.

6.3.1.4 Procedure

All participants completed the experiment using a web based experimental environment developed at Radboud University. During the experiment, participants heard the set only once and indicated which version (LE or GE) of the target word in the set sounded more native to them. The instructions presented on the screen read "Which one sounds more native? Click on Z (first one) or M (second one)." The instructions were shown until the participant made a decision and then there was a one second break before the next set was played. Participants had three self-paced breaks, which were distributed evenly throughout the experiment, each after four speakers. After completing the experiment, participants answered demographic and language background questions. In total, the session lasted approximately 50 minutes.

At the University of Alberta, the experiment was conducted on Think Center Lenovo computers with Windows 7. Between two and 21 participants completed the experiment simultaneously in a room. Participants heard the stimuli through MB-QUART MBK C 800 headphones at a comfortable volume. At UPV/EHU, the experiment was conducted on Mac Mini computers using MacOS Sierra, and stimuli were played over Sennheiser HD 380 Pro model headphones at a comfortable level. Between one and four participants completed the experiment simultaneously in a sound attenuated room. After the completion of the experiment, the Spanish participants additionally completed the LexTALE task (Lemhöfer & Broersma, 2012), where participants indicate whether the visually presented stimulus is a real English word or not. This task estimates the general English proficiency of the individual.

6.3.2 Analyses

We analyzed participants' nativeness judgments with glmers. Our dependent variable was whether the participant indicated the GE or the LE version of the word as sounding more

native (coded as 0 for GE and 1 for LE). The predictors of interest were Target Word Category (/θ/, schwa, and voiced obstruent target words), Speaker Nativeness (native, non-native), and Listener Nativeness (native, non-native). The scale function in R (R Core Team, 2016) was used to center and scale Trial Number, which served as a control predictor to measure practice effect and fatigue. Additionally, Order of concatenation (order 1, order 2) was included as a control predictor as participants may be influenced by whether they hear the GE or LE token first.

We followed the same procedure as with the computer evaluation (Experiment 1) in determining the best model. This meant that we began with interactions among our predictors of interest (Target Word Category, Speaker Nativeness, and Listener Nativeness). The model was stripped until we were left with only significant predictors. The significant predictors were checked as random slopes. For crossed random effects, we had Target Word, Speaker, and Listener. As there was a significant three-way interaction (Target Word Category* Speaker Nativeness* Listener Nativeness), we split the data by Target Word Category in order to better interpret the results. Upon splitting, if the interaction between Speaker and Listener Nativeness was not significant, we removed the interaction. Then, if the predictors themselves, Listener Nativeness and Speaker Nativeness, were not significant, we also removed them from the model. The control variables, which were simple effects, were kept in the split models whether significant or not as they applied to the complete dataset.

In the models reported below, the schwa target word category (Target Word Category), native speaker (Speaker Nativeness), and native listener (Listener Nativeness*)* are on the intercept. We chose the schwa target word category for the intercept as we expected a difference between Spanish and Canadian listeners. We had native listener and native speaker on the intercept since we see them as our baseline.

Due to technical issues, we randomly lost up to 14 trials from each Spanish participant (a total of 154 trials across the 39 participants). Additionally, one Spanish participant only completed part of the experiment due to hardware issues (here the loss was 221 trials). This resulted in a total of 16,902 data points for the Spanish and 19,492 for the Canadian participants.

6.3.3 Results

The final model presented in Table 5 revealed a significant effect of Trial Number as well as Order. The listeners' preference for LE increased as the experiment progressed (significant effect of Trial Number). Additionally, the preference for LE decreased when the sets of target words were presented in order 2 (significant effect of Order).

Table 5: *Table of Type II Wald Chi-square Tests of the glmer model for listeners' responses to the different target words. Significance values (p) are indicated using the following ranges >0.1, <0.1. <0.05, <0.01, and <0.001, where anything <0.05 or below is considered significant.*

Fixed Effects	*Chi-square*	*p*
Target Word Category	2.7	>0.1
Speaker Nativeness	3.1	<0.1
Listener Nativeness	9.6	<0.01
Order	22.3	<0.001
Trial Number	8.9	<0.01
Target Word Category * Speaker Nativeness	6.7	<0.5
Target Word Category * Listener Nativeness	14.5	<0.001
Speaker Nativeness * Listener Nativeness	0.9	>0.1
Target Word Category * Speaker Nativeness * Listener Nativeness	6.8	<0.05

The model also showed a three-way interaction (Target Word Category * Speaker Nativeness * Listener Nativeness). To better understand the data, we split the data by Target Word Category (/θ/, schwa, and voiced obstruent) and analyzed the sub-datasets separately.

6.3.3.1 /θ/ Target Word Category

Listeners' responses to the /θ/ target words are visualized in Figure 2 and the corresponding statistical model is shown in Table 6. The non-significant intercept (p = 0.054) in the model indicates that the native listeners did not have a clear preference for either GE or LE as more native-like when listening to the native speakers. The simple effect of Listener Nativeness indicates that non-native listeners had less of a preference for LE, while it is not clear whether they had a preference for GE. Therefore, in order to better understand the role of Listener Nativeness, we releveled the model to have the non-native listeners on the intercept. When the model was releveled, the intercept remained non-significant (β = -0.36, z = -1.85, p < 0.1) and therefore the non-native listeners did not show a clear preference for GE. In conclusion, both native and non-native listeners did not have a clear preference for LE or GE, whether listening to native or non-native speakers, even though there is a difference between the two speaker groups in their preference.

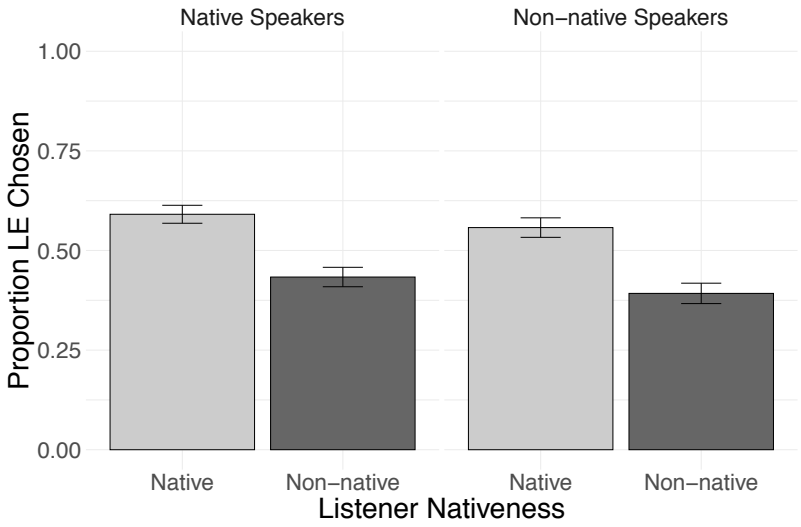

Figure 2: *Listeners' responses to /θ/ target words split by listener and speaker nativeness.*

The random effect structure shows the preference differed per target word, listener and per speaker (significant random intercept of Target Word, of Listener and of Speaker), as well as differing per listener nativeness by speaker (significant random slope of Listener Nativeness by Speaker). Additionally, the order of concatenation affected listeners differently (significant random slope of Order by Listener).

Table 6: *The glmer model for listeners' responses to /θ/ target words with native listeners and native speakers on the intercept. Significance values (p) are indicated using the following ranges >0.1, <0.1. <0.05, <0.01, and <0.001, where anything <0.05 or below is considered significant.*

Fixed Effects	*β*	*z*	*p*
(Intercept)	0.40	1.92	<0.1
Listener Nativeness: Non-Native	-0.76	-3.62	<0.001
Order	-0.12	-1.72	<0.1
Trial Number	0.03	1.46	>0.1
Random Effects			*SD*
Listener (intercept)			0.86
Order by Listener			0.51
Speaker (intercept)			0.52
Listener Nativeness by Speaker			0.40
Target Word (intercept)			0.28

6.3.3.2 Schwa Target Word Category

Listeners' responses to the schwa target words are shown in Figure 3 and the statistical model is presented in Table 7. Overall, the native listeners found the LE realization more native than the GE realization when listening to native speakers, as indicated by the significant intercept. More importantly for our research question, the model shows significant simple effects of Speaker Nativeness and Listener Nativeness and their interaction. If either the speaker or the listener was non-native there was less of a preference for LE, but this was less so if both the speaker and the listener were non-native. To determine whether the simple effects and interaction led to the absence of a preference for LE, we releveled the model and examined the significance of the intercept (see Appendix 2 for the releveled models). We did not find significant intercepts for the releveled models, indicating that only when the native speakers heard native listeners was there a preference for the LE. In the other combinations, where the speaker or listener, or both, were non-native, there was no statistically significant preference for LE (Non-native listener and native speaker on the intercept: $\beta = 0.20$, $z = 0.98$, $p > 0.1$, native listener and non-native speaker: $\beta = 0.33$, $z = 1.51$, $p > 0.1$, non-native listener and non-native speaker: $\beta = -0.16$, $z = -0.80$, $p > 0.1$).

The random effect structure is very similar to the one for the /θ/ target words. It shows that the preference differed per target word, per listener and per speaker (significant random intercepts of Target word, Listener and Speaker), and that the effect of listener nativeness differed by speaker (significant random slope of Listener Nativeness by Speaker). Additionally, the order of concatenation affected listeners differently (significant random slope of Order by Listener).

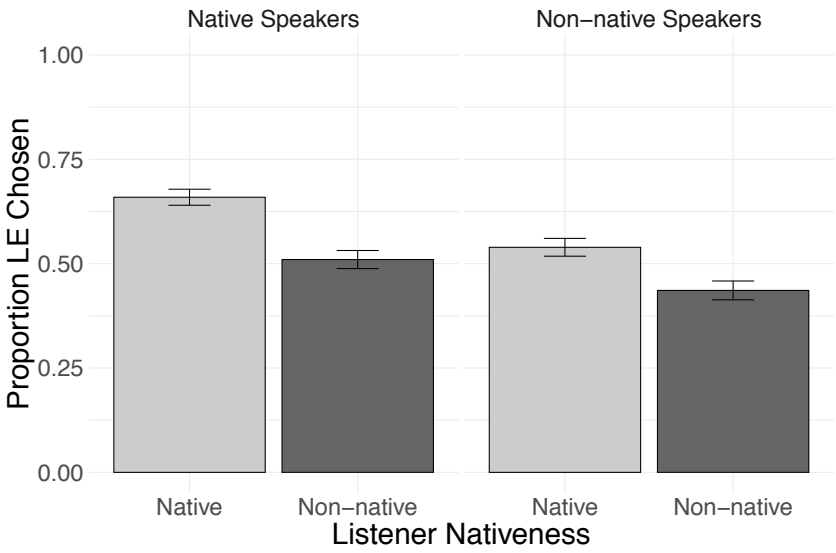

Figure 3: *Listeners' responses to schwa target words split by listener and speaker nativeness.*

Table 7: *The glmer model for listeners' responses to schwa target words with native listeners and native speakers on the intercept. Significance values (p) are indicated using the following ranges >0.1, <0.1. <0.05, <0.01, and <0.001, where anything <0.05 or below is considered significant.*

Fixed Effects	*β*	*z*	*p*
(Intercept)	0.95	4.36	<0.001
Speaker Nativeness: Non-native	-0.62	-2.93	<0.01
Listener Nativeness: Non-native	-0.75	-3.23	<0.01
Order	-0.37	-5.00	<0.001
Trial number	0.05	2.76	<0.01
Speaker Nativeness : Non-native * Listener Nativeness: Non-native	0.26	2.29	<0.05
Random Effects			*SD*
Listener (intercept)			1.02
Order by Listener			0.56
Speaker (intercept)			0.41
Listener Nativeness by Speaker			0.17
Target word (intercept)			0.13

6.3.3.3 Voiced Obstruent Target Word Category

Listeners' responses to the voiced obstruent target words are illustrated in Figure 4 and the statistical model can be seen in Table 8. The significant intercept indicates that native listeners again preferred LE as the most native sounding token. This preference seems independent of whether they listened to native or non-native speech, as Speaker Nativeness was not statistically significant. The significant simple effect of Listener Nativeness means that this preference for LE was smaller for non-native listeners. The model was releveled to examine whether the preference for LE for the non-native listeners was statistically significant (non-native listener on the intercept). This appeared not to be the case ($\beta = 0.05$, $z = 0.26$, $p > 0.1$).

The random structure was the same as for the other two target word categories. Again, the listeners' preference for LE differed in size per target word, speaker and listener (as shown by significant random intercepts for Target Word, Speaker and Listener), the order of concatenation affected listeners differently (significant random slope of Order by Listener), and the random effect of speaker differed by listener's nativeness (significant random slope of Listener Nativeness by Speaker).

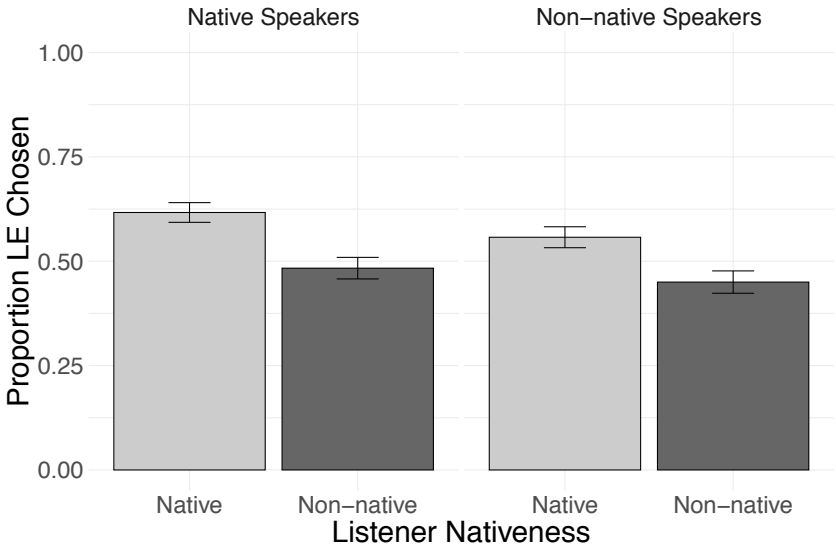

Figure 4: *Listeners' responses to voiced obstruent target words split by listener and speaker nativeness.*

Table 8: *The glmer model for listeners' responses to voiced obstruent target words with native listeners and native speakers on the intercept. Significance values (p) are indicated using the following ranges >0.1, <0.1. <0.05, <0.01, and <0.001, where anything <0.05 or below is considered significant.*

Fixed Effects	*β*	*z*	*p*
(Intercept)	0.61	3.40	<0.001
Listener Nativeness: Non-native	-0.56	-2.81	<0.01
Order	-0.38	-5.17	<0.001
Trial Number	0.02	0.74	>0.1
Random Effects			*SD*
Listener (intercept)			0.92
Order by Listener			0.53
Speaker (intercept)			0.39
Listener Nativeness by Speaker			0.31
Target Word (intercept)			0.16

6.4 Discussion

In this article, we investigated *within-speaker* variation in the degree of the non-native accent as modulated by the effort speakers invest in their pronunciation. We compared target words that were produced in two extreme conditions of articulatory effort: one produced in noise with contrastive focus on the target words and the other in quiet without contrastive focus on the target words. In the former condition, we expected more articulatory effort (Greater Effort - GE condition) than in the latter condition (Less Effort - LE condition). We analyzed the speech signal using a forced speech aligner (Computer Evaluation: Experiment 1) as well as by having human listeners evaluate the accentedness of the target words (Human Evaluation: Experiment 2). This combination of computer and human evaluation provides a more holistic understanding of the accentedness of the target words. While the computer evaluation provided an objective evaluation of the presence of the phone of interest, the human listeners evaluated the whole word, including other phones in the word, speech rate, and suprasegmental features such as intonation. Moreover, the listeners were likely to be influenced by their native languages.

The target words were produced by native American-English and non-native – native Dutch – speakers of English in the two distinct articulatory effort conditions, GE and LE. We chose three target word categories that pose difficulties for native Dutch speakers in English and that vary in Dutch speakers' awareness of the difficulty. These categories consisted of /θ/-initial words, Dutch-English cognates with schwa in pre-stress position in English and a full vowel in its place in the corresponding Dutch cognates, and words with final voiced obstruents (hereafter referred to as /θ/, schwa, and voiced obstruent target word category, respectively).

We hypothesized that the effect of articulatory effort is different for the three types of phonemes. Since Dutch speakers are aware of the difficulty of /θ/, we expected that increased effort (GE condition) would improve their pronunciation. On the other hand, for the schwa target words, we expected that, since vowels are more often reduced in non-focused position in Dutch, the LE condition would encourage more productions of schwa, and thus more native like pronunciations. As for the voiced obstruent category, we did not expect the Dutch speakers to change their pronunciation in either GE or LE. Since they are not aware that they have final devoicing in their native language without realizing it, they would also likely have it in their non-native language.

We used the forced speech aligner MFA (McAuliffe et al., 2017) to examine the phone of interest in each of these target words. In the forced alignment process, MFA could choose from the American-English and the Dutch-accented English pronunciations of the target words, which differed in the transcription of the phone of interest. As expected, MFA's

transcriptions showed that for all three target word categories, the non-native speakers produced fewer prototypical American-English phones of interest compared to the native speakers (/θ/: non-natives 55.13% and natives 92.36%, schwa: non-natives 70.1% and natives 84.6%, and voiced obstruent: non-natives 56.3% and natives 85.2%).

More importantly, the schwa target word category revealed a difference between the GE and LE condition. The non-native speakers produced more schwas in the LE condition (79.3%) than the GE condition (60.9%). While this is in line with our predictions for the non-native speakers, unexpectedly, we found that the native speakers showed a similar difference in schwa production in the LE (94.3%) and GE (75%) conditions. A likely explanation is that not only the non-native but also the native speakers' production of schwas was influenced by the spelling of the target words (the vowel was orthographically written as <a> and <o>), especially when they tried to articulate more clearly in the GE condition.

While there was a difference between GE and LE in the production of the prototypical phone of interest for the schwa target word category, we did not find this for the /θ/ and voiced obstruent categories. We had expected this absence of a difference for voiced obstruents, because most Dutch speakers are unaware of the phenomenon of final devoicing in their language and therefore cannot easily repair it. In contrast, we had expected a statistically significant difference between the greater and less effort conditions (GE and LE) for /θ/, whose difficulty Dutch speakers are aware of. There may be several reasons for why we did not find this difference. First, producing speech in noise may be so cognitively demanding for non-native speakers, that they may not have further resources left to improve their pronunciation of this phone. Second, the non-native speakers may simply not be able to improve its pronunciation. Third, the non-native speakers may have been subtly altering their pronunciations, but this was not detected by MFA since the differences were not large enough to span a whole phone boundary. Future research has to further investigate these reasons.

In short, we only observed an effect of articulatory effort for classification of schwa target words, and this difference was the same for the native and in the non-native speakers. Possibly, we would have found our hypothesized effects (i.e. differential effects of effort on the native and the non-native speakers) if we had increased the statistical power by analyzing all 30 non-native speakers in the DELNN corpus instead of the eight mid-accented ones. However, it is also very likely that the hypothesized effects differ with the speaker's proficiency. For instance, the effect of articulatory effort on the pronunciation of /θ/ is likely to be absent in both low proficiency speakers (as they simply cannot pronounce the phone correctly) and in high proficiency speakers (as they always produce the phone correctly) and only be present in mid-accented speakers. Analyzing the whole corpus would therefore not only imply analyzing more data, but also having a larger proportion of the data that are likely not to show effects.

For the human listeners' evaluation, native (Canadian) and non-native (native Spanish) listeners of English heard the target words by the same speaker produced in the two articulatory effort conditions (GE and LE) and indicated which one sounded more native like. We had native and non-native listeners, expecting that their accentedness evaluation would differ due to their familiarity with the target phonemes from their native language. We specifically expected differences for the schwa category because English has schwa in its phonemic inventory, while Spanish does not. The two languages are similar in that both Spanish and English have /θ/ in their phonemic inventory.

In analyzing human listeners' evaluations we found that native listeners had a preference for the LE condition as sounding more native like when listening to native speakers produce the schwa and voiced obstruent words. The preference for LE was expected, considering that all speech was presented to listeners in quiet, and Lombard speech (GE) heard in quiet may sound unnatural since it is presented outside of its natural (noisy) context.

That we found the difference for schwa words with native listeners is in line with the MFA results, but not for the words with voiced obstruents, for which the MFA did not observe a difference in pronunciation between the LE and GE conditions. The difference in the MFA results between the schwa and the voiced obstruents words correlates with the size of the effect of articulation effort on accentedness as perceived by the native listeners (which is greater for the schwa than for the voiced obstruent words). That the native listeners also perceived an effect of articulation effort on the voiced obstruent words, while our MFA analysis did not, may be because articulatory effort affected phones other than those we examined in the MFA analysis or the effects may have been too subtle to be detected by MFA.

We did not find a statistically significant preference for LE for /θ/ words, although there was a trend for a preference for LE. Moreover, while the native and non-native listeners differed in their preference, the non-natives did not have a statistically significant preference either. The lack of a clear preference for LE for the /θ/ words may be due to the differences being very subtle for /θ/ words or to the fact that /θ/ is word-initial while the other consonants under consideration, namely voiced obstruents, eliciting a preference, was more prominent as word-final.

When the native listeners heard non-native speech, they mostly showed the same preferences as for native speech. The native listeners were not judging the non-native speakers differently than the native speakers. There is one exception, however. For schwa words produced by non-native speakers, the native listeners did not show any preference, neither for the words produced in LE nor those produced in GE. This is striking because it is only for these words that the MFA determined that the non-native speakers produced more phones that are prototypical English in the LE condition than in the GE condition. Possibly,

in the LE condition, these words were produced with other non-native like characteristics than the pretonic vowel tested in the MFA analysis, and these other characteristics may be dominant for the native listeners (e.g., the amount of aspiration on the preceding stop).

What our results for the native listeners therefore show is that if non-native speakers put more effort in their speech (GE) for the voiced obstruent words, they are judged as sounding less native like. Articulating clearly may be one of the mechanisms that can make speech sound non-native like. This could be because the level of clarity does not match the context, as is the case for native Lombard speech presented in quiet, or because non-natives' attempt at clear articulation is less native like their than non-clear articulation. This is in line with Oyama's research (1976), who found that increased attention had a negative effect on pronunciation if the speaker did not speak English well.

This finding shows social consequences of speaking clearly for non-native speakers. A whole body of literature (e.g., Fuse et al., 2018; Hendriks et al., 2018; Kalin & Rayko, 1978; Roessel et al., 2019) has shown that listeners evaluate speakers more negatively when they have stronger non-native accents. By knowing when an individual's accent may be the least accented, the individual could use this information to their advantage to decrease the accent and in turn dampen the negative judgments associated with it. Our results suggest that if non-native speakers put more effort in their speech for the voiced obstruent words, this may have negative consequences for how native listeners judge them. If non-native speakers may feel the need to articulate as clearly as possible in a quiet environment, they may not want to clarify as if they were in a noisy environment with contrastive focus on the target word, because then the clarification strategy may have negative social consequences.

Non-native listeners did not show any preferences for either the LE or the GE speech for any of the word categories. It is likely that non-native listeners are used to more pronunciation variation than native listeners because they more often hear non-native speech, from different speakers with varying proficiency levels, than native listeners. Therefore, non-native listeners may be more accepting and flexible in their judgments of nativeness. Furthermore, non-native listeners may be used to foreign directed speech (FDS), which shares features with Lombard speech (GE condition was Lombard speech with contrastive focus on the target word), so they do not have as strong a preference for LE as native listeners. Among other changes, both FDS and Lombard speech have a slower speech rate, undergo changes in the vowel space, and have longer vowels (for a review of the acoustic modifications associated with FDS and Lombard speech see: Cooke, King, Garnier, & Aubanel, 2014).

In conclusion, we documented how articulatory effort as induced by noise and contrastive focus affects the degree of non-native accent when presented in quiet. We compared native and non-native – native Dutch – speakers of English. Non-native listeners hardly vary their

accentedness judgements as a function of the speaker's articulatory effort. This is different for native listeners: they prefer pronunciations produced with less articulatory effort for many words. With the exception of schwa words produced by native speakers, this preference is not reflected in the words' transcriptions focusing on the phonemes that are problematic for Dutch speakers of English. The preference must therefore stem from more subtle deviations. Our results suggest that for most words it does not matter whether non-native speakers produce them with more articulatory effort. For some words increased articulatory effort makes the non-native speakers sound less native-like, which may come with negative social implications attached to a non-native accent.

Chapter 7:
General discussion

This dissertation examined the characteristics of non-native Lombard speech using three separate approaches; investigating acoustics, intelligibility, and accentedness. Combined, these approaches give insight into both the production and perception of non-native Lombard speech. I will discuss how the results of these approaches fit together. I focused on Lombard speech produced by native speakers of Dutch, by comparing the differences between Lombard and plain speech in their non-native English with their native Dutch and with native speakers of English.

7.1 DELNN Corpus

I created the DELNN (Dutch English Lombard Native Non-Native) corpus as a first step in addressing the research questions on the characteristics of non-native Lombard speech. The corpus consists of a total of 39 female speakers; 30 Dutch natives producing both native Dutch and non-native English plain and Lombard speech, and nine American-English speakers producing native English plain and Lombard speech. Speakers read question-answer pairs. The English pairs each contained a target word that was selected because of a phonological characteristic. The target words formed three categories, each one posing different difficulties for the non-native speakers; 1.) /θ/-initial words, 2.) English-Dutch cognates with schwa in the pre-stress position in the English pronunciation and a full vowel in its place in the Dutch pronunciation, and 3.) words with final voiced obstruents. Target words were embedded in all the answers, and in half of the answers, they received contrastive focus (referred to as the late-focus condition). In the other half of the answers, a word earlier in the answer, not the target word, received contrastive focus (referred to as the early-focus condition). The answers of the Dutch question-answer pairs also varied in the early versus late position of contrastive focus. There are a total of 144 English and 96 Dutch question-answer pairs, half of which were produced as plain and the other half as Lombard speech. To my knowledge, the DELNN corpus is the first corpus to include non-native Lombard speech that has been made publicly available for researchers.

In addition to the recordings discussed, the corpus also includes accompanying textgrids aligning the acoustic signal of the answers with phone- and word-level annotations, which also reflect false starts and word substitutions for the answers. The textgrids were produced by adapting existing forced aligners. The Dutch textgrids were created using Kaldi (Povey et al., 2011) and the English textgrids were created using the Montreal Forced Aligner, which

uses Kaldi as its basis (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017). For this dissertation, the corpus was used for acoustic analyses and material was extracted as stimuli for intelligibility and accentedness experiments.

In having this research be one of the first steps towards investigating non-native Lombard speech, I wanted to choose two languages that had similarities as well as differences. Moreover, I wanted non-native speakers with an adequately high proficiency in English. This is how I came to choose English and Dutch as the two languages in the DELNN corpus. In general, the Dutch are known to be quite proficient in English, making them a good choice as speakers for the corpus. Additionally, there are segmental and prosodic differences between the Dutch and English languages, which make the combination interesting to study. The main differences I was interested in were at the phoneme level, more specifically, the lack of /θ/ in the Dutch phonemic inventory, the presence of final devoicing in Dutch (final voiced obstruents are devoiced), differences in voice onset time (VOT) for stop consonants, and in the frequency of schwa in English and Dutch.

While I chose to investigate the combination of Dutch and English, future research could examine more distinct language combinations. This may show a greater influence of the native language in the non-native Lombard speech in the acoustic analyses, which would likely influence the perception of the non-native Lombard speech. For instance, investigating two languages, one being a tonal language and the other not, may further reveal native language influence on Lombard speech.

As discussed above, Dutch speakers were in part chosen because of their high proficiency in English. When recruiting participants, I specifically looked for Dutch individuals who were attending university (which meant they pertained to a limited age group) and were not completing their degree in English. I was cautious that speakers who were completing their degree in English may be too proficient and may consequently show no effect of the native language on the non-native language. Even with this requirement in place, some of the participants were highly proficient, as revealed by the range in the evaluation of speaker's accentedness (as judged by native English listeners in an accentedness evaluation). It is possible that if non-native speakers with lower proficiencies were recorded, I would observe greater influences of the native language, especially in such acoustic measures as VOT length.

The DELNN corpus only contains female speakers. Females were chosen since some studies have shown that they differ from male speakers in their pitch and energy modifications (Junqua, 1993) as well as in vowel duration and F2 changes (Alghamdi, Maddock, Marxer, Barker, & Brown, 2018) in their Lombard speech. Furthermore from the perception side, Junqua (1993) found female Lombard speech to be more intelligible than male Lombard speech. Moreover, some acoustic measurements such as fundamental frequency (median, and

range) differ between males and females, therefore only having female speakers simplifies the analysis. Recruiting female participants is also easier. Future research should show to what extent differences between native and non-native Lombard speakers differs between female and male speakers. I would expect the overall pattern of results to be similar.

It should be noted that while there is scarce research on individual differences in Lombard speech, it has been suggested that these differences exist (e.g., Junqua, 1993; Pittman & Wiley, 2001, Shen, 2022). While I examined group differences (native versus non-native) in this dissertation, the DELNN corpus could also be used to examine individual differences. Having a homogeneous group of speakers (female, university students, with a specific age, and with a specific English proficiency) minimizes the diversity in the group and allows one to examine how individuals differ in their Lombard speech production.

An important feature of the DELNN corpus is that the recordings consist of read speech. I made this decision to be able to use the DELNN corpus for acoustic analyses and to extract stimuli for perception experiments. By having read speech, I guaranteed that the content of the recordings were similar across participants and importantly, across native and non-native speakers. This decision reduced variability in the materials and allows for more direct comparisons.

While I believe that using read speech as a basis for the DELNN corpus was appropriate, I also believe that, as Tucker and Ernestus (2016) argue, investigating spontaneous non-native Lombard speech is worthwhile. Casual speech is inherently more ecologically valid, and differs from read speech in many respects. Considering that casual speech has a greater communicative purpose than read speech and that communicative purpose has been shown to increase the extent of acoustic modifications of Lombard speech (e.g., Junqua, Fincke, & Field, 1999; Lane & Tranel, 1971; Villegas, Perkins, & Wilson, 2021), one could possibly expect larger acoustic effects in casual Lombard speech. However, read speech gives us more control over the output, taking away variability from context (e.g., word duration being affected by predictability, i.e. context) that would occur in spontaneous speech, which makes it easier to isolate the effects of Lombard speech as opposed to other effects. Therefore, as a next step, it would be insightful to examine casual non-native Lombard speech.

7.2 Acoustics

7.2.1 Acoustic Measures

Regarding acoustics, I was interested in examining whether there were potential differences between native and non-native English Lombard speech and in turn whether these potential differences were influences of the native language or perhaps effects of speaking a non-native language. In Chapters 2, 3, and 4, I examined acoustic modifications that are well documented

in native Lombard speech. I investigated fundamental frequency (F0), F0 range, spectral Center of Gravity (CoG), intensity, and duration (for a review of acoustic modifications of Lombard speech see: Cooke et al., 2014). I further decided to explore VOT as it has not been studied in Lombard speech (no article to my knowledge) and because I expected greater native language influence on this specific acoustic measure compared to the other measures. In choosing the acoustic measures, I wanted to have a range from global to more fine grained, looking at the sentence (median F0, F0 range, spectral CoG, and intensity), word (duration), and phone (VOT) level.

For each acoustic measure, I had two analyses: one comparing native and non-native English speech (English speech, between-speakers) and one comparing native Dutch and non-native English speech (Dutch speakers, within-speakers). The first analysis, examining native and non-native English, allowed me to investigate how the non-native English speakers adapt their Lombard speech in comparison to the native speakers. The second comparison, between native Dutch and non-native English, examines if the speakers adapt their speech similarly in noise in their native and non-native language, potentially explaining differences found in the first analysis. Combined, these two analyses allowed for a better understanding of non-native Lombard speech.

A handful of articles has examined a set of acoustic measures (sound pressure levels, intensity, mean F0 of the vowel, mean intensity of the vowel, and vowel duration) in non-native English Lombard speech produced by Chinese and Japanese native speakers (Cai, Yin, & Zhang, 2020, 2021; Mok, Li, Luo, & Li, 2018; Villegas, Perkins, & Wilson, , 2021). My research adds to this past research not only in studying a different pair of languages but also by studying several acoustic measures not yet documented for non-native Lombard speech. In line with this past research, the non-native speakers in the DELNN corpus increased their intensity and word duration in Lombard speech. Further, I found that non-native speakers in these studies widened their F0 ranges and showed increases for median F0 and spectral CoG in Lombard speech, which is newly documented in non-native Lombard speech research but in line with past research on native Lombard speech. New to Lombard speech research is that I found that VOT length decreased in Lombard speech as compared to plain speech for both native and non-native speakers.

The comparisons between the native English and non-native English speakers on the one hand and between the Dutch native speakers when speaking Dutch and English on the other hand suggest that there is an effect of the native language on the non-native Lombard speech modifications for some acoustic measures. The presence of this influence in the data appeared to be modulated by the position of contrastive focus in the answer of the question-answer pair. Thus, the native Dutch speakers showed a similar increase in median F0 from plain to Lombard speech in their native Dutch and their non-native English for early- and late-

focus. With respect to the English speech, the native and non-native English speakers showed a similar increase in F0 in the late-focus condition, while for the early-focus condition I only found a significant increase in F0 for the non-native English speakers. Regarding the change in F0 range, for late-focus the size of the F0 range increase was the largest for the native English speakers, followed by the non-native English speakers, and then by the same speakers in their native Dutch. When investigating median F0, I decided to analyze F0 using hertz (Hz) rather than semitones because I was more interested in production rather than perception. To verify whether the interaction of nativeness and Lombard speech remained, I converted the data from Hz to semitones and re-analyzed it. I found that while the detailed statistics slightly changed, the main pattern of results was the same. Combined, I interpret these median F0 and F0 range results to indicate an influence of Dutch on non-native English Lombard speech.

In addition, there were acoustic measures that differed similarly between plain and Lombard speech in native English and non-native English. This was the case for spectral CoG, intensity, duration, and VOT, in which the non-native English speakers adapted their Lombard speech similarly to native English speakers. This indicates that, depending on the acoustic measure under investigation, the native and non-native English speakers may be adapting their Lombard speech similarly.

Simultaneously, for some acoustic measures I found differences in Lombard speech between the native Dutch and the non-native English from the same speakers. This was the case for spectral CoG, intensity, and VOT. The native Dutch speech revealed a larger increase in spectral CoG, a smaller increase in intensity for late-focus, and a smaller decrease in VOT when going from plain to Lombard speech in Dutch than in their non-native English. This would suggest that the native Dutch speakers are adapting their Lombard speech differently in the two languages for certain acoustic measures.

The one acoustic measure that did not reveal any differences between plain and Lombard speech across the native English, non-native English, and native Dutch speech, was word duration.

These acoustic results combined indicate that, overall, the non-natives are adapting their speech in noise. For some acoustic properties, non-native speakers showed exactly the same adaptations as native English speakers (intensity, spectral CoG, VOT, word duration). For these measures the non-native speakers had the same adaptations in their native language (word duration) or different adaptions in their native language (intensity, spectral CoG, VOT). For other acoustic properties (median F0 and F0 range), non-natives speakers showed the influence of their native language in their adaptations for Lombard speech. It remains an open question why some acoustic measures showed an influence of the native language

while others do not. My research suggests that it is not the level (sentence, word, phone) that explains these differences. Further research should also investigate whether the absence of differences may be due to the fact that the learners have learnt to adapt their non-native Lombard speech or because the two native languages have some phonological or phonetic differences that prompt subtle differences in some acoustic adaptations in Lombard speech.

In addition, future research could analyze the same acoustics, such as spectral CoG and duration, at the phoneme level, which may lead to more nuanced analyses, further documenting a potential influence of the native language on the non-native Lombard speech. Analyzing the acoustic measures at different levels is especially important considering that vowels and consonants may differ in their modifications, as is the case with vowels being elongated more in Lombard speech than consonants (e.g., Castellanos, Benedí, & Casacuberta, 1996; Garnier & Henrich, 2014; Junqua, 1993) and differences in the amount spectral CoG increases for various phones for Lombard speech relative to plain speech (e.g., Junqua, 1993; Lu & Cooke, 2008).

Finally, in the analysis of median F0, it appeared that there is less creaky voice in Lombard speech as compared to plain speech. This intuition was not examined further as the scope of this dissertation did not allow for it, however future work may be interested in examining this further.

7.2.2 Acoustics: Computational measures

While Chapters 2, 3, and 4 analyzed several traditional acoustic measures, I further investigated acoustics in parts of Chapters 5 and 6 using computational measures. One computational measure was the High Energy Glimpsing Proportion metric (HEGP; Tang & Cooke, 2016), which indicates the speech's ability to withstand noise, in part by measuring the energy in the speech. I used the HEGPs to analyze a subset of English target words from eight native and eight non-native speakers in the DELNN corpus. The HEGPs revealed that native and non-native English Lombard speech had higher HEGPs than plain speech. These findings for native speech are in line with previous research (using GPs, the predecessor to HEGPs: Lu & Cooke, 2009b). Importantly, the results from the HEGPS of non-native speech are a new finding, in line with my direct acoustic measures, documenting non-natives' Lombard speech production. Further, I found that native English speech had higher HEGPs than non-native English speech. As HEGPs indirectly examine the energy in the speech to determine its ability to withstand noise, this would suggest that the energy in the speech of native and non-natives may vary. It is unclear why this is the case and should be investigated further.

Another computational measure was the results of the forced aligner, which provided phone-level transcriptions of the target words. The forced aligner was allowed to choose

from the prototypical English transcription of the word or a Dutch pronunciation. Therefore, from this transcription, I was able to determine per the forced aligner whether the phones of interest in the target words were pronounced in a prototypical English manner or not. The forced aligner examined two conditions in the DELNN corpus, which maximally differed in their articulatory effort: Lombard speech with contrastive focus on the target word (where I expected greater effort, GE) and plain speech with the target word not receiving contrastive focus (where I expected less effort, LE). The results revealed that, while the non-native speakers had fewer instances of prototypical English phones compared to the native speakers, the difference between the two conditions was similar for the native and non-native speech – only showing changes in pronunciation for the schwa target word category. For the schwa target words, both natives and non-native speakers had fewer prototypical instances of schwa in GE than LE. This indicates that while non-native speakers produce fewer instances of prototypical phones than native English speakers, they parallel the change that native speakers make between the GE and LE conditions.

The different patterns observed for the different acoustic and computational measures suggests that Lombard speech is complex and that only examining a few acoustic measures for the influence of the native language is not enough. This dissertation therefore has to be considered as one of the first steps in investigating non-native Lombard speech. It is important that future research examines more acoustic measures at various levels (e.g., vowel space, spectral CoG at the phone level).

7.3 Intelligibility

Past research with native and non-native speakers listening to native speech has shown that Lombard speech is more intelligible than plain speech when both are heard in noise (native listeners: e.g., Dreher & O'Neill, 1957; Lu & Cooke, 2008; Pittman & Wiley, 2001, non-native listeners: e.g., Cooke & García Lecumberri, 2012). The purpose of Chapter 5 was to study the size of the Lombard intelligibility benefit for native and non-native English speech. Because we know that a speaker's intelligibility may not only be determined by the acoustic signal but also by the listener's language background, I tested native Canadian listeners and two non-native listener groups, Dutch and Spanish. The Dutch listeners shared their native language with the non-native speakers while the Spanish listeners did not. In plain speech, Bent and Bradlow (2003) have shown that if the non-native listener shares the native language with the non-native speaker, they can find them as intelligible as native speakers, having a matched interlanguage speech intelligibility benefit. If there are language specific modifications of Lombard speech, one may expect to see a difference between the two non-native listener groups, with the Dutch listeners having a larger Lombard intelligibility benefit

than the Spanish listeners for the non-native speech due to them sharing the native language of the speakers.

The results from the intelligibility experiment revealed that both native and non-native speech elicited a Lombard intelligibility benefit. While past research has shown native speech to elicit a Lombard benefit (e.g., Dreher & O'Neill, 1957; Lu & Cooke, 2008; Pittman & Wiley, 2001), to my knowledge, this is the first study that has investigated and found a Lombard benefit for non-native speech. Finding a Lombard benefit for non-native speech strongly suggests that non-native speakers produce Lombard modifications that listeners find beneficial when listening in noise. Of note, I did not find a difference in the size of the Lombard benefit for native and non-native speech. This suggests, that any acoustic differences between the modifications in native and non-native Lombard speech that I found in Chapters 2-4 do not impact its intelligibility, or at least not significantly.

In the study, for native and non-native speech combined, the non-native listeners experienced a slightly larger Lombard intelligibility benefit (16.5%) than the native listeners (13.4%). In contrast, I did not find a difference in the size of the Lombard benefit between the two non-native listener groups. It could be the case that the languages I chose – Dutch and Spanish – are too similar, and in order to see the effects of the language specific modifications in the Lombard intelligibility benefit, I would need different language combinations. This is a topic that future research should address.

Performing this research led to some additional interesting findings not discussed in the content chapters, as they are not directly related to the main research question. These are discussed further below.

The intelligibility experiment indicated that the two non-native listener groups overall differed from each other, with the Dutch performing better in general than the Spanish. This difference could appear for several reasons, including a difference in proficiency. While I did not find a statistically significant difference between the Dutch and Spanish in their English proficiency at the group level, as indicated by their LexTALE scores (Lemhöfer & Broersma, 2012), there was a numerical difference, and I therefore investigated whether the difference between the Dutch and Spanish listeners remained if I took each listener's individual English proficiency into account. I ran a separate analysis (not part of Chapter 5) on the Dutch and Spanish listeners' intelligibility scores, excluding the native English listeners, and included their LexTALE scores as a predictor of interest in the model. This analysis revealed that even when their LexTALE scores are included in the model, the Dutch and Spanish listeners differed from each other, with the Dutch listeners again having higher intelligibility scores overall.

Because the difference between the two non-native listener groups remains when

individual proficiency is included, something more inherent to the listeners themselves may be driving the difference between the groups. One possible explanation could have to do with the characteristics of the native language (Dutch and Spanish) and in turn the similarities and differences between Dutch, on the one hand, and Spanish, on the other hand, with English. One example is that the Spanish phonemic inventory does not contain schwa while the Dutch and the English ones do.

I tested the intelligibility of plain and Lombard speech by mixing it with masking noise, at a fixed signal to noise ratio (SNR) of -1 dB. Possibly, the finding that the native listeners showed a smaller Lombard intelligibility benefit than the non-native listeners could potentially be due to native listeners performing at ceiling when listening to native Lombard speech. This could be investigated by determining the size of a Lombard benefit using the speech reception threshold (SRT) method. The SRT establishes the SNR required to achieve a certain percentage (e.g., 50%) of correctly identified speech. While fixed SNRs have typically been used to investigate a Lombard intelligibility benefit (e.g., Lu & Cooke, 2008; Pittman & Wiley, 2001), the SRT technique has been used to measure native and non-native speech intelligibility (e.g., van Wijngaarden, 2001). In line with past research on the Lombard benefit, I decided to use fixed SNRs. I tested one fixed SNR. A follow-up experiment could test an additional (lower) fixed SNR level (e.g., SNR of -5 dB), which would allow one to explore whether the native listeners were performing at ceiling level when listening to native Lombard speech. Alternatively, one could use the SRT method with the data to examine this matter further.

7.4 Accentedness

In Chapter 6, I investigated the perception of non-native accent as affected by articulatory effort (vocal and articulatory), which is an important feature of Lombard speech. I examined listeners' accentedness evaluations of native and non-native speech, presenting target word tokens produced by the same speaker as Lombard speech with contrastive focus and as plain speech without contrastive focus. I expected greater effort (GE) for the former condition and less effort (LE) for the latter. Participants indicated which of the two realizations by the same speaker was more native-like. The target words belonged to the three categories incorporated in the DELNN corpus (/θ/-initial words, English-Dutch cognates with schwa in the pre-stress position in English and a full vowel in Dutch, and words with final voiced obstruents). I had both native (Canadian) and non-native (native Spanish) English listeners for the accentedness evaluation, as the listeners may be influenced by their native language and therefore differ in their evaluation of native and non-native speech. Spanish listeners were chosen since Spanish shares some features with English that distinguish it from Dutch, namely that the Spanish

phonemic inventory includes /θ/. The Spanish language differs from English in that schwa is not in its phonemic inventory.

I expected that listeners would prefer the LE realizations because Lombard speech (GE target words) heard in quiet may sound unnatural, as it is not heard in the noisy environment for which it was produced. Nevertheless, the non-native listeners did not show a preference for either realization, GE or LE, when listening to native or non-native speech. The adaptations coming with Lombard speech thus did not affect the preference, neither in native nor in non-native speech, for the non-native listeners. It could be that the non-native listeners did not show a preference because the adaptations associated with GE and LE did not make a difference for them. Alternatively, as non-native listeners, they may be flexible in their evaluations.

This was different for the native listeners. As expected, the native listeners tended to have a preference for the LE realizations of target words with schwa or voiced final obstruents but I did not find this preference for /θ/ target words when listening to native speech. The absence of a preference is in line with the forced aligner results, which similarly transcribed /θ/ in the /θ/ target words in GE and LE.

When listening to non-native speech, native listeners only showed the preference for LE realizations for voiced final obstruent words. Hence, native speakers' evaluations of native and non-native speech was mostly the same, except for the schwa target word category. This difference may either be driven by differences in the acoustic modifications implemented in schwa words between native and non-native speakers (as indicated by the forced aligner), or by the fact that for native listeners the non-native schwa words were too accented for the Lombard modifications to make much of a difference.

Considering that for the most part speaker nativeness did not affect the preference between LE and GE for each listener group, this may suggest that the non-native speakers are making similar acoustic Lombard modifications to the native speakers. This would be in line with the findings from Chapters 2-4, where native and non-native English speakers made similar modifications in Lombard speech for all acoustic measures, while differing in the extent of the modifications for some. Further, the accentedness results roughly parallel the intelligibility findings in Chapter 5, where I did not find differences in the size of the Lombard benefit for native and non-native speech.

I had Canadian listeners as the native group and Spanish listeners as the non-native group. It would have been informative to additionally have Dutch listeners. That way I could have had a group of non-native listeners who shared their native language with the non-native speakers. I therefore recommend future research to investigate other groups of non-native listeners, especially Dutch listeners.

As native and non-native listeners perceive non-native accents more negatively in terms of the speaker's personality and competencies (e.g., Fuse, Navichkova, & Alloggio, 2018; Hendriks, van Meurs, & Hogervorst, 2016; Hendriks, van Meurs, & Reimer, 2018; Tsurutani, 2012), this article sheds light on the accentedness evaluation of Lombard speech and its societal consequence. Considering that the native listeners in this study tended (for two of the three target word categories) to have the same preference or lack thereof for native and non-native speech, this suggests that whatever native language influences may be present in the non-native Lombard speech, they may not be so drastically large as to alter the evaluation of non-native speech compared to native speech.

While for most words it does not matter whether non-native speakers produce them with much or little effort, for some words much articulatory effort makes them sound less native-like for native listeners. Therefore, there may be reason to recommend non-native speakers to change the amount of effort they put into their pronunciation.

7.5 Conclusion

In using the DELNN corpus to examine non-native Lombard speech from three distinct perspectives – acoustics, intelligibility and accentedness –, I can conclude that non-native speakers produce Lombard speech. While I found subtle influences of the native language on non-native Lombard speech, this was only the case for a couple of the several acoustic measures analyzed. These subtle influences of the native language do not seem to affect the perception of Lombard speech, as I found similar sized Lombard intelligibility benefits for native and non-native speech. Further, in terms of accentedness, the two articulatory effort conditions were rated mostly similarly for native and non-native speech, indicating that the native and non-native speakers are behaving similarly.

This first exhaustive study of non-native Lombard speech thus shows that native Dutch speakers produce non-native English Lombard speech that may not be completely identical to native English Lombard speech, but is highly similar.

References

Alghamdi, N., Maddock, S., Marxer, R., Barker, J., & Brown, G. J. (2018). A corpus of audio-visual Lombard speech with frontal and profile views. *The Journal of the Acoustical Society of America*, *143*(6), EL523–EL529. https://doi.org/10.1121/1.5042758

Asher, J. J., & García, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal*, *53*(5), 334–341. https://doi.org/10.1111/j.1540-4781.1969.tb04603.x

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)* (Release 2, Dutch Version 3.1) [Data set]. Linguistic Data Consortium.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–1111. https://doi.org/10.1016/j.jml.2008.06.003

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Berendsen, E. (1986). The phonology of Dutch cliticization. In W. de Gruyter (Ed.), *The Phonology of Cliticization* (pp. 35–98). Foris Publications.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 171–204). New York Press.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer (Version 6.0.37) [Computer software]. http://www.praat.org/

Booij, G. (1985). Lexical phonology, final devoicing and subject pronouns in Dutch. *Linguistics in the Netherlands*, 21–26. https://doi.org/10.1515/9783112330128-005

Booij, G. (1999). *The phonology of Dutch.* Oxford University Press.

Bosker, H. R., & Cooke, M. (2018). Talkers produce more pronounced amplitude modulations when speaking in noise. *The Journal of the Acoustical Society of America*, *143*(2), EL121–EL126. https://doi.org/10.1121/1.5024404

Bosker, H. R., & Cooke, M. (2020). Enhanced amplitude modulations contribute to the Lombard intelligibility benefit: Evidence from the Nijmegen Corpus of Lombard Speech. *The Journal of the Acoustical Society of America*, *147*(2), 721–730. https://doi.org/10.1121/10.0000646

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Burgos, P., Cucchiarini, C., van Hout, R., & Strik, H. (2013). Pronunciation errors by Spanish learners of Dutch: A data-driven study for ASR-based pronunciation training. In F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, and P. Perrier (Eds.) *Proceedings of the 14th Annual Conference of the International Speech Communication Association (INTERSPEECH 2013)* (pp. 2385–2389).

Cai, X., Yin, Y., & Zhang, Q. (2020). A cross-language study on feedforward and feedback control of voice intensity in Chinese–English bilinguals. *Applied Psycholinguistics*, *41*(4), 771–795. https://doi.org/10.1017/S0142716420000223

Cai, X., Yin, Y., & Zhang, Q. (2021). Online control of voice intensity in late bilinguals' First and second language speech production: Evidence from unexpected and brief noise masking. *Journal of Speech, Language, and Hearing Research*, *64*(5), 1471–1489. https://doi.org/10.1044/2021_JSLHR-20-00330

Campbell, W. N. (1995). Loudness, spectral tilt, and perceived prominence in dialogues. In K. Elenius, & Branderud (Eds.) *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 676–679).

Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. In A. Botinis (Ed.) *Proceedings of Intonation: Theory, Models and Applications*.

Castellanos, A., Benedí, J. M., & Casacuberta, F. (1996). An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication*, *20*(1–2), 23–35. https://doi.org/10.1016/S0167-6393(96)00042-8

Chen, Y. (2015). Post-fous compression in English by Mandarin learners. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. The University of Glasgow.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33(2)*, 121–157. https://doi.org/10.1016/j.wocn.2005.01.001

Choi, H. (2003). Prosody-induced acoustic variation in English stop consonants. In M. J. Solé, D. Recasens, and J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. (pp. 2661–2664).

CMU Pronouncing Dictionary (2015). (Version 0.7b) [Data set]. http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics*, *35*(2), 180–209. https://doi.org/10.1016/J.WOCN.2006.03.004

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

Cooke, M., & García Lecumberri, M. L. (2012). The intelligibility of Lombard speech for non-native listeners. *The Journal of the Acoustical Society of America*, *132*(2), 1120–1129. https://doi.org/10.1121/1.4732062

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 & Language*, *28*(2), 543–571. https://doi.org/10.1016/j.csl.2013.08.003

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America*, *128*(4), 2059–2069. https://doi.org/10.1121/1.3478775

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142–2156. https://doi.org/10.1121/1.392372

Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division (Strasbourg). (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Press Syndicate of the University of Cambridge.

Cutler, A. (2012). Second-language listening: Sounds to words. In *Native listening: Language experience and the recognition of spoken words* (pp. 303–335). MIT Press.

Dreher, J. J., & O'Neill, J. (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *The Journal of the Acoustical Society of America*, *29*(12), 1320–1323. https://doi.org/10.1121/1.1908780

Dutch Language Institute (2014). *Corpus Gesproken Nederlands - CGN* (Version 2.0.3) [Data set]. http://hdl.handle.net/10032/tm-a2-k6

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1–3), 162–173. https://doi.org/10.1006/brln.2001.2514

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, *76*(3), 692–707. https://doi.org/10.1121/1.391256

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 233–277). New York Press.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185–202. https://doi.org/10.1016/0167-6393(87)90025-2

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134. https://doi.org/10.1121/1.413041

Flege, J. E., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English/t/-/d/contrast. *Journal of Phonetics*, *17*(4), 299–315. https://doi.org/10.1016/S0095-4470(19)30446-2

Fónagy, I., & Fónagy, J. (1966). Sound pressure level and duration. *Phonetica*, *15*(1), 14–21. https://doi.org/10.1159/000258534

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*(5), 489–504. https://doi.org/10.1016/0749-596X(87)90136-7

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third Ed.). Sage Publications.

Fuse, A., Navichkova, Y., & Alloggio, K. (2018). Perception of intelligibility and qualities of non-native accented speakers. *Journal of Communication Disorders*, *71*, 37–51. https://doi.org/10.1016/J.JCOMDIS.2017.12.006

García Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. https://doi.org/10.1016/J.SPECOM.2010.08.014

Garnier, M. (2008). May speech modifications in noise contribute to enhance audio-visible cues to segment perception? In R. Göcke, P. Lucey, & S. Lucey (Eds.), *Proceedings of International Conference on Auditory-Visual Speech Processing 2008* (pp. 95–100). International Speech Communication Association Archive.

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Loevenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. In *Proceedings of the Ninth International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP)*.

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. https://doi.org/10.1016/J.CSL.2013.07.005

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. https://doi.org/10.1016/j.csl.2013.09.007

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice*, *2*(2), 118–126. https://doi.org/10.1016/S0892-1997(88)80067-5

Gussenhoven, C., & Broeders, A. (1997). Intonation. In *English Pronunciation for Student Teachers* (Second Ed, pp. 174–188). Wolters-Noordhoff.

Gustafson, E., Engstler, C., & Goldrick, M. (2013). Phonetic processing of non-native speech in semantic vs non-semantic tasks. *The Journal of the Acoustical Society of America*, *134*(6), EL506–EL512. https://doi.org/10.1121/1.4826914

Hanssen, J. E. G., Peters, J., & Gussenhoven, C. (2008). Prosodic effects of focus in Dutch declaratives. In Plínio A. Barbosa, Sandra Madureira, and Cesar Reis (Eds.), *Proceedings of Speech Prosody 2008* (pp. 609–612).

Hanulikova, A., & Weber, A. (2010). Production of English interdental fricatives by Dutch, German, and English speakers. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 173–178). Adam Mickiewicz University.

Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of English revisited: Do non-native listeners feel the same? *Research in Language*, *15*(3), 285–298. https://doi.org/10.1515/rela-2017-0016

Hazan, V., Grynpas, J., & Baker, R. (2012). Is clear speech tailored to counter the effect of specific adverse listening conditions? *The Journal of the Acoustical Society of America*, *132*(5), EL371–EL377. https://doi.org/10.1121/1.4757698

Hendriks, B., van Meurs, F., & Hogervorst, N. (2016). Effects of degree of accentedness in lecturers' Dutch-English pronunciation on Dutch students' attitudes and perceptions of comprehensibility. *Dutch Journal of Applied Linguistics*, *5*(1), 1–17. https://doi.org/10.1075/dujal.5.1.01hen

Hendriks, B., van Meurs, F., & Reimer, A. K. (2018). The evaluation of lecturers' nonnative-accented English: Dutch and German students' evaluations of different degrees of Dutch-accented and German-accented English of lecturers in higher education. *Journal of English for Academic Purposes*, *34*, 28–45. https://doi.org/10.1016/J.JEAP.2018.03.001

Hualde, J., Olarrea, A., Escobar, A., & Travis, C. (2001). *Introducción a la lingüística hispánica*. Cambridge University Press.

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, *117*(2), 896–907. https://doi.org/10.1121/1.1823291

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Junqua, J. C. (1994). A duration study of speech vowels produced in noise. In *Proceedings of the Third International Conference on Spoken Language Processing* (pp. 419–422).

Junqua, J. C., Fincke, S., & Field, K. (1999). The Lombard effect: A reflex to better communicate with others in noise. In *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP99 (Cat. No. 99CH36258)* (pp. 2083–2086). IEEE.

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. https://doi.org/10.1121/1.405631

Kager, R. W. J. (1989). Secondary stress and vowel reduction in Dutch. In *A Metrical Theory of Stress and Destressing in English and Dutch* (pp. 275–31). Utrecht University.

Kalin, R., & Rayko, D. S. (1978). Discrimination in evaluative judgments against foreign-accented job candidates. *Psychological Reports*, *43*(3_suppl), 1203–1209. https://doi.org/10.2466/pr0.1978.43.3f.1203

Keating, P. A., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 2–7). The University of Glasgow.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, *24*(4), 372–392. https://doi.org/10.1016/S0163-6383(02)00086-3

Kormos, J. (2006). Monitoring. In *Speech Production and Second Language Acquisition* (pp. 122–136). Lawrence Erlbaum Associates.

Lacabex, E. G., García Lecumberri, M., & Cooke, M. (2005). English vowel reduction by untrained Spanish learners: Perception and production. In *Proceedings of Phonetics Teaching & Learning Conference*.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research* , *14*(4), 677–709. https://doi.org/10.1044/jshr.1404.677

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/J.JESP.2010.05.025

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In *Speech Production and Speech Modelling* (pp. 403–439). Springer. https://doi.org/10.1007/978-94-009-2037-8_16

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*(1), 1–28. https://doi.org/10.1177/002383096701000101

Lombard, E. (1911). Le signe de l'elevation de la voix (The sign of the elevation of the voice). *Ann. Mal. de L'Oreille et Du Larynx*, *37*, 101–119.

Lotz, J., Abramson, A. S., Gerstman, L. J., Ingemann, F., & Nemser, W. J. (1960). The perception of English stops by speakers of English, Spanish. Hungarian, and

Thai: A tape-cutting experiment. *Language and Speech*, *3*(2), 71–77. https://doi.org/10.1177/002383096000300202

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

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

Lu, Y., & Cooke, M. (2009b). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication*, *51*(12), 1253–1262. https://doi.org/10.1016/j.specom.2009.07.002

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, *36*(2), 173–190. https://doi.org/10.2307/3588329

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2017)* (pp. 498–502). https://doi.org/10.21437/interspeech.2017-1386

Mennen, I., Schaeffler, F., & Docherty, G. (2007). Pitching it differently: A comparison of the pitch ranges of German and English speakers. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1769–1772).

Missaglia, F. (1999). Contrastive prosody in SLA: An empirical study with adult Italian learners of German. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 551–554).

Mixdorff, H., Pech, U., Davis, C., & Kim, J. (2007). Map Task dialogs in noise–a paradigm for examining Lombard speech. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1329–1332).

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331–340. https://doi.org/10.3758/BF03211209

Mok, P., Li, X., Luo, J., & Li, G. (2018). L1 and L2 phonetic reduction in quiet and noisy environments. In *Proceedings of the 9th International Conference on Speech Prosody 2018* (pp. 848–852). https://doi.org/10.21437/SpeechProsody.2018-171

Mulac, A., & Rudd, M. J. (1977). Effects of selected American regional dialects upon regional audience members. *Communication Monographs*, *44*(3), 185–195. https://doi.org/10.1080/03637757709390130

Munro, M. J., Derwing, T. M., & Sato, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers. *Prospect*, *21*(1), 67–79.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261–283. https://doi.org/10.1007/BF01067377

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). Librispeech: An ASR corpus based on public domain audio books. In *Proceedings of* the 2015 *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 5206–5210). IEEE.

Perkell, J. S., Denny, M., Lane, H., Guenther, F., Matthies, M. L., Tiede, M., … Burton, E. (2007). Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users. *The Journal of the Acoustical Society of America*, *121*(1), 505-518. https://doi.org/10.1121/1.2384848

Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R., & Kearney, J. K. (1989). Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America*, *85*(2), 894–900. https://doi.org/10.1121/1.397561

Pinet, M., Iverson, P., & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker–listener accent similarity. *The Journal of the Acoustical Society of America*, *130*(3), 1653–1662. https://doi.org/10.1121/1.3613698

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215. https://doi.org/10.1006/JPHO.2001.0134

Pisoni, D., Bernacki, R., Nusbaum, H., & Yuchtman, M. (1985). Some acoustic-phonetic correlates of speech produced in noise. In *Proceedings of ICASSP '85. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol 10, pp. 1581–1584). IEEE. https://doi.org/10.1109/icassp.1985.1168217

Pittman, A. L., & Wiley, T. L. (2001). Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research*, *44*(3), 487–496. https://doi.org/10.1044/1092-4388(2001/038)

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., … Vesely., K. (2011). The Kaldi speech recognition toolkit. In *Proceedings of IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society.

Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*(2), 271–287. https://doi.org/10.1111/j.1467-1770.1980.tb00319.x

Quené, H., Orr, R., & van Leeuwen, D. (2017). Phonetic similarity of /s/ in native and second language: Individual differences in learning curves. *The Journal of the Acoustical Society of America*, *142*(6), EL519–EL524. https://doi.org/10.1121/1.5013149

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, *51*(4B), 1296–1303. https://doi.org/10.1121/1.1912974

Rasier, L., Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, 28(2007), 41-66.

Roessel, J., Schoel, C., Zimmermann, R., & Stahlberg, D. (2019). Shedding new light on the evaluation of accented speakers: Basic mechanisms behind nonnative listeners' evaluations of nonnative accented job candidates. *Journal of Language and Social Psychology*, *38*(1), 3–32. https://doi.org/10.1177/0261927X17747904

Rump, H. H., & Collier, R. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech*, *39*(1), 1–17. https://doi.org/10.1177/002383099603900101

Scharenborg, O., & van Os, M. (2019). Why listening in background noise is harder in a non-native language than in a native language: A review. *Speech Communication*, *108*, 53–64. https://doi.org/10.1016/j.specom.2019.03.001

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America*, *85*(1), 295–312. https://doi.org/10.1121/1.397737

Segalowitz, N. (2010). Second Language Cognitive Fluency. In *Cognitive Bases of Second Language Fluency* (pp. 74–106). Routledge.

Shen, C. (2022). *Indivdual differences in speech production and maximum speech performance.* [Doctoral dissertation , Radboud University].

Simon, E. (2010). Phonological transfer of voicing and devoicing rules: Evidence from L1 Dutch and L2 English conversational speech. *Language Sciences*, *32*(1), 63–86. https://doi.org/10.1016/J.LANGSCI.2008.10.001

Simon, E., & Leuschner, T. (2010). Laryngeal systems in Dutch, English, and German: A contrastive phonological study on second and third language acquisition. *Journal of Germanic Linguistics*, *22*(4), 403–424. https://doi.org/10.1017/S1470542710000127

Simonet, M., Casillas, J. V., & Díaz, Y. (2014). The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/,/t/) and linguistic experience (monolinguals, bilinguals). In *Proceedings of the 7th International Conference on Speech Prosody* (pp. 202–206).

Sityaev, D., & House, R. (2003). Phonetic and phonological correlates of broad, narrow and contrastive focus in English. In M. J. Solé, D. Recasens, and J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1819–1822).

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236–264. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Stibbard, R. M., & Lee, J. I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America*, *120*(1), 433–442. https://doi.org/10.1121/1.2203595

Stoicheff, M. (1981). Speaking fundamental frequency characteristics of nonsmoking female adults. *Journal of Speech, Language, and Hearing Research*, *24*(3), 437–441. https://doi.org/10.1044/jshr.2403.437

Tang, Y., & Cooke, M. (2016). Glimpse-based metrics for predicting speech intelligibility in additive noise conditions. In *Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPECH 2016)* (pp. 2488–2492). https://doi.org/10.21437/Interspeech.2016-14

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*(2), 177–204. https://doi.org/10.1111/j.1467-1770.1991.tb00683.x

Tsurutani, C. (2012). Evaluation of speakers with foreign-accented speech in Japan: The effect of accent produced by English native speakers. *Journal of Multilingual and Multicultural Development*, *33*(6), 589–603. https://doi.org/10.1080/01434632.2012.697465

Tucker, B. V., Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon, 11*(3), 375–400. https://doi.org/10.1075/ml.11.3.03tuc

van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*, *12*(1), 1–23. https://doi.org/10.1016/0167-6393(93)90015-D

van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*(3), 253–265. https://doi.org/10.1177/002383099503800303

van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa. *Language Learning*, *66*(1), 124–158. https://doi.org/10.1111/lang.12141

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, *84*(3), 917–928. https://doi.org/10.1121/1.396660

van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication*, *35*(1–2), 103–113. https://doi.org/10.1016/S0167-6393(00)00098-4

Varadarajan, V. S., & Hansen, J. H. L. (2006). Analysis of Lombard effect under different types and levels of noise with application to in-set speaker ID systems. In *Proceedings of the Ninth International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP)*.

Villegas, J., Perkins, J., & Wilson, I. (2021). Effects of task and language nativeness on the Lombard effect and on its onset and offset timing. *The Journal of the Acoustical Society of America*, *149*(3), 1855–1865. https://doi.org/10.1121/10.0003772

Voigt, R., Jurafsky, D., & Sumner, M. (2016). Between-and within-speaker effects of bilingualism on F0 variation. In *Proceedings of Interspeech 2016: The 17th Annual Conference of the International Speech Communication Association* (pp. 1122–1126).

Welby, P. (2006). Intonational differences in Lombard speech: Looking beyond F0 range. In *Proceedings of the Third International Conference on Speech Prosody* (pp. 763–766).

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

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 152–155).

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Jounral of Phonetics*, *33*(2), 159–197. https://doi.org/10.1016/j.wocn.2004.11.001

Yao, Y. (2009). Understanding VOT variation in spontaneous speech. *UC Berkley PhonLab Annual Report*, *5*(5), 29–43.

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, *37*(2), 231–247. https://doi.org/10.1016/j.wocn.2009.03.002

Zollinger, S. A., & Brumm, H. (2011). The Lombard effect. *Current Biology*, *21*(16), R614–R615. https://doi.org/10.1016/j.cub.2011.06.003

Appendix

Chapter 2:

Appendix 1

Table 1: *English target words used in the DELNN corpus per target word category.*

Schwa	Voiced Obstruent	/θ/
Balloon	Blood	Theater
Banana	Cab	Theme
Botanical	Club	Theology
Cadaver	Crib	Theory
Computer	Food	Therapist
Gorilla	Lab	Thermal
Massage	Lemonade	Thermodynamics
Parade	Neighborhood	Thermometer
Police	Pub	Thermos
Professor	Rehab	Theta
Tomato	Road	Thriller
Salami	Wood	Throne

Appendix 2

Table 2: *Dutch target words used with their English translation.*

Dutch target word	English translation
Ballon	Balloon
Kadaver	Cadaver
Computer	Computer
Gorilla	Gorilla
Banaan	Banana
Massage	Massage
Politie	Police
Professor	Professor
Tomaat	Tomato
Botanische	Botanical
Salami	Salami
Parade	Parade
Universiteit	University
Hoofdgerecht	Main course
Appartement	Apartment
Bibliotheek	Library
Kostuum	Costume
Telefoon	Telephone
Oorbellen	Earrings
Museum	Museum
Artikel	Article
Programma	Program
Rugzak	Backpack
Gladiool	Gladiolus

Appendix 3: MFA Evaluation

The English data were annotated at the phone and word level with the Montreal Forced Aligner (MFA; McAuliffe et al., 2017), which uses Kaldi as its basis (Povey et al., 2011). In order to apply forced alignment, the speech signal (WAV files), an orthographic transcription, a pronunciation dictionary, and the phone models were provided. In the orthographic transcription of what the speaker produced, false starts were included for the answers so that the best phonetic annotation possible could be provided. The dictionary provides a map from the words (in the orthographic transcription) to the phones.

Since the English stimuli were recorded by native English as well as native Dutch speakers, Dutch-accented pronunciations of the target words were included in the pronunciation dictionary. We used the Carnegie Mellon University (CMU) Pronouncing Dictionary (*CMU Pronouncing Dictionary*, 2015), which has the American-English pronunciations of words and added Dutch-accented variants for the three target word categories. For the /θ/-initial target words, /t/, /d/, /f/, /v/, /s/, and /z/ were included as alternative pronunciations of /θ/. As for the schwa target words, alternative pronunciations were included where schwa was replaced by /ʌ/, /æ/, /ɑ/, and /ɔ/ if the schwa was orthographically spelled as <a>, and replaced by /ɔ/, /o/, and /ɑ/ if it was spelled with <o>. For the target words with final voiced obstruents, we included variants with /t/ and /p/ at the end of the word instead of /d/ and /b/, respectively.

The acoustic models used for the transcription of the English utterances were English phones trained on the LibriSpeech corpus, with 1000 hours of read speech (Panayotov et al., 2015). First, MFA calculated monophone Gaussian Mixture Models-based (GMM) Hidden Markov Models (HMMs) and then, in order to take the surrounding phones into account, calculated triphone GMM-HMM models (McAuliffe et al., 2017). MFA calculated 13 mel-frequency cepstral coefficients (MFCCs), as well as 13 each for delta and delta-delta, resulting in 39 features per frame. The acoustic models resulted in 68 total GMM-HMMs (one for each vowel and consonant, and additional ones for e.g. silences). Cepstral mean and variance normalization (CMVN) was applied per speaker. Speaker adaptation was not implemented since it did not improve phone-level transcription.

In order to evaluate the phone level transcriptions for the English utterances, we had two trained human annotators annotate 25 sentences from 13 non-native English speakers, of which 14 sentences were plain speech and 11 were Lombard speech. We found that overall, the MFA annotation was comparable to that produced by the human annotators (see the all phones section in Table 3).

Since silences influence spectral CoG and intensity values at the sentence level, we needed to remove them and therefore we specifically examined how well MFA annotated silences. The agreement between each of the human transcriptions and the MFA transcriptions was overall lower than the agreement between the two human transcriptions, but this was especially so for the silences' boundaries. In order to investigate this further, the first author annotated silences from 60 randomly selected utterances not considered in the evaluation (a combination of plain and Lombard speech as well as a combination of native English and non-native English), resulting in manual annotation of 117 silence start boundaries and 55 silence end boundaries (there were more start boundaries than end boundaries because many end boundaries were sentence final and therefore not annotated). While the annotations of the silence start boundaries did not show a consistent pattern and could therefore not be improved, these annotations suggested that the MFA silence end boundaries were on average 30 ms late, unless they were sentence final. We therefore lengthened the non-sentence final silences by 25 ms (5 ms less than the average difference in order to ensure that only in few cases the annotated silence lasted in the next phone). This change improved the agreement between the MFA and human annotators, increasing the number of silence end boundaries that were within 25 ms of each other.

Table 3 illustrates the agreement of the MFA transcription with each of the two human transcribers and the agreement between the two human transcribers themselves. It shows the effect of lengthening the non-utterance final silence boundaries by 25 ms. A 25 ms window was chosen as to be able to compare MFA performance's on our data to its evaluation of other corpora (see McAuliffe et al., 2017). Table 3 indicates that after lengthening the non-utterance final silence boundaries, of the phones with the same labels, 75.4% and 75.1% were less than 25 ms off from Human 1 and Human 2's annotations, respectively. These figures are in the same range as the forced aligners trained by other researchers, for instance, McAuliffe and colleagues (2017) found that 77% of the aligned phone boundaries were less than 25 ms off from the gold-standard annotations for the Buckeye corpus and 72% for the Phonsay corpus.

	Pre-Move			Post-Move		
	All labels	Same labels (%)	Boundaries within 25 ms (%)	All labels	Same labels (%)	Boundaries within 25 ms (%)
Silence start boundary						
MFA-Human 1	56	39 (69.6)	25 (64.1)	56	39 (69.6)	25 (64.1)
MFA-Human 2	51	40 (78.4)	25 (62.5)	51	40 (78.4)	25 (62.5)
Human 1-Human 2	47	37 (78.7)	33 (89.2)	47	37 (78.7)	33 (89.2)
Silence end boundary						
MFA-Human 1	56	27 (48.2)	6 (22.2)	57	27 (47.4)	21 (77.8)
MFA-Human 2	51	28 (54.9)	8 (28.6)	52	28 (53.9)	22 (78.6)
Human 1-Human 2	47	37 (78.7)	36 (97.3)	47	37 (78.7)	36 (97.3)
All phones						
MFA-Human 1	803	712 (88.7)	525 (73.7)	804	710 (88.3)	535 (75.4)
MFA-Human 2	798	645 (80.8)	476 (73.8)	799	643 (80.5)	483 (75.1)
Human 1-Human 2	792	676 (85.4)	593 (87.7)	792	676 (85.4)	593 (87.7)

Table 3: *Statistics on label and boundary agreement between MFA, Human 1 and Human 2, for pre- and post-moving of the silence end boundary. 'All labels' is the combined number of phones that the two annotators transcribed for the specified category (silence, or all phones). 'Same labels (%)' is the number of phones for which the two annotators agreed upon the labeling followed by its percentage of the 'All labels', in parenthesis. The 'Boundaries within 25 ms (%)' is the number of boundaries that the two annotators labeled the same (Same labels) and were within 25 ms of each other, followed by the percentage in parenthesis.*

The Dutch data were directly annotated with Kaldi (Povey et al., 2011) since there were no acoustic models for Dutch built into MFA. As with the English annotation, the orthographic transcription of what the speaker produced included false starts for the answers. The pronunciation dictionary was created from a combination of Celex (Baayen, Piepenbrock, & Gulikers, 1995) and the Spoken Dutch Corpus (CGN; Dutch Language Institute, 2014). The acoustic models were trained on the complete CGN except for the telephone recordings, which have a low acoustic quality. As was the case with MFA (McAuliffe et al., 2017), Kaldi (Povey et al., 2011) computed 13 MFCCs and the delta and delta-delta, for a total of 39 features per frame. The training resulted in 50 nnet3 triphone models (DNNs) for vowels, consonants and other speech sounds. Cepstral mean and variance normalization (CMVN) was applied per utterance.

In order to evaluate Kaldi's (Povey et al., 2011) transcription, two different human annotators annotated 25 Dutch utterances, 15 plain utterances from a separate corpus and 10 Lombard utterances from DELNN. The results are presented in Table 4. As with the English transcriptions, we found that the annotations of the Dutch silence boundaries could be improved. In order to calculate how the silence boundaries should be adjusted for improvement, the first author annotated 30 answers (20 plain and 10 Lombard) from the DELNN corpus not included in the evaluation. This led to a total of 86 silence start and 57 silence end boundaries (excluding utterance final boundaries) annotations. These annotations indicated that the silence start boundary should be moved forward by 20 ms and the silence end boundary should be lengthened by 20 ms. These changes improved the transcription, as can be seen in Table 4 below. This resulted in Human 1 and Kaldi having 82.2% of all phones (that had the same label) within 25 ms of each other and 86.6% for Human 2 and Kaldi. Here we see that the evaluation of the Dutch transcriptions which used Kaldi, is even better than the evaluation of the English transcriptions above for MFA, which is similar to the evaluation of the forced aligner by McAuliffe and colleagues (2017).

	Pre-Move			Post-Move		
	All labels	Same labels (%)	Boundaries within 25 ms (%)	All labels	Same labels (%)	Boundaries within 25 ms (%)
Silence start boundary						
Kaldi-Human 1	44	36 (81.8)	17 (47.2)	44	36 (81.8)	24 (66.7)
Kaldi-Human 2	38	30 (79.0)	15 (50.0)	38	30 (79.0)	24 (80.0)
Human 1-Human 2	42	30 (71.4)	22 (73.3)	42	30 (71.4)	22 (73.3)
Silence end boundary						
Kaldi-Human 1	44	26 (59.1)	3 (11.5)	44	26 (59.1)	11 (42.3)
Kaldi-Human 2	38	23 (60.5)	4 (17.4)	38	23 (60.5)	18 (78.3)
Human 1-Human 2	42	33 (78.6)	25 (75.8)	42	33 (78.6)	25 (75.8)
All phones						
Kaldi-Human 1	1045	885 (84.7)	711 (80.3)	1045	885 (84.7)	727 (82.2)
Kaldi-Human 2	1041	860 (82.6)	723 (84.1)	1041	860 (82.6)	745 (86.6)
Human 1-Human 2	988	893 (90.4)	818 (91.6)	988	893 (90.4)	818 (91.6)

Table 4: *Statistics on label and boundary agreement between Kaldi (Povey et al., 2011), Human 1 and Human 2, for pre- and post-moving of the silence start and end boundaries. 'All labels' is the combined number of phones that the two annotators transcribed for the specified category (silence, or all phones). 'Same labels (%)' is the number of phones which the two annotators agreed upon the labeling, followed by its percentage of the 'All labels', in parenthesis. The 'Boundaries within 25 ms (%)' is the number of boundaries that the two annotators labeled the same ('Same labels') and were within 25ms of each other, followed by the percentage in parenthesis.*

Appendix 4

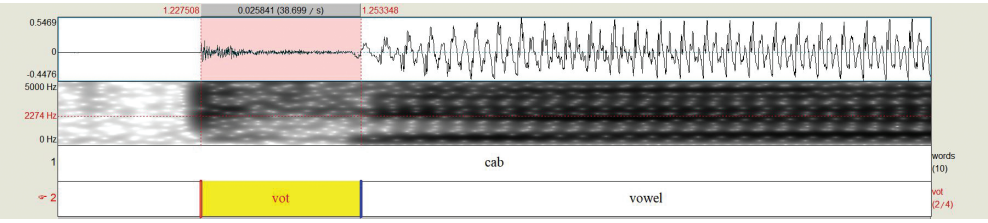

Figure 1: *Example of a VOT annotation.*

Appendix 5: VOT: Interrater Agreement

There were three annotators, of which one annotated tokens of all the target words while each of the two others only annotated part of the target words, not overlapping. Human 1 and Human 2 both annotated English *computer*, *cadaver*, *parade*, *professor*, *police*, and *cab*. To determine the interrater agreement, we had a total of 60 annotations from each annotator, nine per target word except for *pub,* which had five. Human 2 and Human 3 both annotated the English *crib* and *club*, and the Dutch *kadaver*, *kostuum*, *computer*, *parade*, and *politie*. For determining the interrater agreement, we had a total of 62 annotations from each annotator, nine per target word except for *crib*, which had eight. Table 5 below shows the results of this comparison

Table 5: *The average and standard deviation of the difference between the two human annotators for VOT start boundary, VOT end boundary and vowel end boundary.*

	Human 1- Human 2		Human 2 – Human 3	
Boundary	*M*	*sd*	*M*	*sd*
VOT start	-1.3 ms	11.5 ms	0.0 ms	1.6 ms
VOT end	0.2 ms	2.8 ms	-1.8 ms	2.1 ms
Vowel end	0.4 ms	15.0 ms	-6.1 ms	12.7 ms

Chapter 5:

Appendix 1

Table 1: *List of the target words.*

Baby	Flower	Neighborhood	Sundays
Balloon	Food	Night	Table
Banana	Force	Notebook	Teacher
Beach	Foundation	Pants	Theater
Birthday	French	Parade	Theme
Board	Fridays	Party	Theology
Bonfire	Game	Pizza	Theory
Boy	Garden	Police	Therapist
Cadaver	Gloves	Professor	Thriller
Café	Gorilla	Rain	Throne
City	Guests	Road	Time
Classes	House	Room	Today
Club	Jazz	Salami	Tomato
Computer	July	Sample	Tomorrow
Conference	Left	Sandwiches	Town
Counselor	Lemonade	Saturday	Tubes
Dance	Letter	Snack	Walk
Day	Likes	Spanish	Watch
Desk	Literature	Spring	Week
Detail	Meeting	Square	Wild
Dinner	Minutes	Store	Woman
Drink	Month	Street	Wood
Education	Morning	Summer	Year
Fall	Move	Sun	Zoo

Chapter 6:

Appendix 1

Table 1: *Target words used per target word category and their Dutch translations.*

θ		**Schwa**		**Voiced Obstruent**	
Theater	Theater	Balloon	Ballon	Blood	Bloed
Theology	Theologie	Banana	Banaan	Club	Club
Theory	Theorie	Botanical	Botanisch	Crib	Wieg
Therapist	Therapeut	Cadaver	Kadaver	Lemonade	Lemonade
Thermal	Thermisch	Computer	Computer	Neighborhood	Buurt
Thermos	Thermosfles	Gorilla	Gorilla	Rehab	Rehabilitatie
Theta	Theta	Parade	Parade	Road	Weg
Thriller	Thriller	Police	Politie	Wood	Hout
Throne	Troon	Professor	Professor		
		Salami	Salami		
		Tomato	Tomaat		

Appendix 2: Human evaluation of the schwa target word category

	NS&NL			NS&NNL			NNS&NL			NNS&NNL		
Fixed Effects	*β*	*z*	*p*	*β*	*z*	*p*	*β*	*z*	*p*	*β*	*z*	*p*
(Intercept)	0.95	4.36	<0.001	0.20	0.98	>0.1	0.33	1.51	>0.1	-0.16	-0.80	>0.1
Speaker Nativeness	-0.62	-2.93	<0.01	-0.36	-2.19	<0.05	0.62	2.93	<0.01	0.36	2.19	<0.05
Listener Nativeness	-0.75	-3.23	<0.01	0.75	3.23	<0.01	-0.49	-2.12	<0.05	0.49	2.13	<0.05
Order: 2	-0.37	-5.00	<0.001	-0.37	-5.00	<0.001	-0.37	-5.00	<0.001	-0.37	-5.00	<0.001
Trial number	0.05	2.76	<0.01	0.05	2.76	<0.01	0.05	2.76	<0.01	0.05	2.76	<0.01
Speaker Nativeness * Listener Nativeness	0.26	2.29	<0.05	-0.26	-2.29	<0.05	-0.26	-2.29	<0.05	0.26	2.29	<0.05
Random Effects			*SD*			*SD*			*SD*			*SD*
Listener (intercept)			1.02			1.02			1.02			1.02
Order by Listener			0.56			0.56			0.56			0.56
Speaker (intercept)			0.41			0.31			0.41			0.31
Listener Nativeness by Speaker			0.17			0.17			0.17			0.17
Target word (intercept)			0.13			0.13			0.13			0.13

Table 2: *The glmer model for listeners' responses to schwa target words. The different columns show the releveled models and their intercept (NS-native speaker, NL-native listener, NNS-non-native speaker, NNL-non-native listener). Significance values (p) are indicated using the following ranges >0.1, <0.1. <0.05, <0.01, and <0.001, where anything <0.05 or below is considered significant.*

Data Management Plan

Personal data:

I collected and/or processed the following personal data: speech recordings, history of hearing or speech problems, and audiograms. Further, I collected information on age, gender, education background, language background, use of (multiple) languages, which while not personal data individually, when combined may be identifiable.

These data contain special categories of personal data: history of hearing or speech problems

It was necessary to collect or process these personal data to achieve the goals of my project because I need to acoustically analyze the speech recordings and use the speech recordings for perception experiments. The background information (language background, use of (multiple) languages, and audiogram) and the history of hearing and speaking problem questions were used to exclude participants from analyses if they did not meet our requirements.

I ensured that I did not collect more personal data than necessary for achieving the goals of my research project by only collecting the minimal personal data used to exclude participants.

I am going to retain the personal data that was needed to answer my research questions for the following 10 years.

Protection of participants' privacy:

Upon data collection, each participant was given a participant number from which point forward they were referred to by. The informed consent forms were scanned and encrypted so as to be securely stored. These encrypted informed consent forms are stored in a different location than the rest of the data. The physical consent forms from data collected in Nijmegen, The Netherlands and Vitoria, Spain were given to the Lab Manager to be stored securely in the CLS lab for six months before they were disposed of. For the participants who participated in the speech recordings, the Data Steward and I have a key linking the consent forms to the participant codes which will be kept as long as the audio recordings are stored for possible re-use. The Canadian consent forms were kept in Canada in line with their ethics policy, and therefore I do not have access to these forms (physical or electronic).

As I used voice recordings, the data cannot be fully anonymized. Only those participants that gave explicit informed consent were included in the DELNN corpus which is stored on Zenodo. The corpus can be distributed to researchers once they sign a form indicating the data will only be used for research purposes and they will adhere to the regulations in the

document, which is also signed by the head of CLS on behalf of the Dean of the Faculty of Arts at Radboud University.

Data storage in the context of scientific integrity:

I followed the policy of my institute and archived the research data associated with my publication (including raw data, metadata and documentation) in a folder in my Radboud University work group folder (i.e., in my "werkgroepmap") for a minimum of 10 years. Access to this folder is restricted to me, my supervisor and the data steward of the faculty.

Public availability in the context of data reuse:

I have made the DELNN corpus available via Zenodo (https://zenodo.org/record/4267819#.YfPj3P7MI2w, doi: 10.5281/zenodo.4267819). This includes the speech recordings of 39 participants as well as the textgrid files with word- and phone-level transcriptions. A codebook explaining the acronyms, a mapping of participants to the lists they received (the list of the randomized stimuli), and orthographic transcriptions of what the speakers actually said (including false starts) for the answers of the question-answer pairs is also included in the data available on Zenodo. The corpus can be distributed to researchers once they sign a form indicating the data will only be used for research purposes and they will adhere to the regulations in the document, which is also signed by the head of CLS on behalf of the Dean of the Faculty of Arts at Radboud University.

Further, I have made my data public via RIS for the articles which have been published and plan to do the same for future articles when they are published based on this data. For Chapters 3, 4, and 5, which have been published, I put the following on RIS: data file used in analysis, a codebook describing the variables in the datafile, and a document about the methodology.

- Chapter 3: https://doi.org/10.17026/dans-xyq-27cd
- Chapter 4: https://doi.org/10.17026/dans-zzn-5tu9
- Chapter 5: https://doi.org/10.17026/dans-xnd-7jfd

Nederlandse samenvatting

Stel je voor dat de werkdag voorbij is, en je bent met je vrienden in een bar. Je moet dan spraak verstaan in een lawaaiige omgeving. Spraak die geproduceerd wordt terwijl er veel achtergrondgeluid is, ook wel Lombard spraak genoemd, klinkt anders dan spraak die geproduceerd wordt in een stille omgeving (normale spraak). Mijn onderzoek gaat over de Lombard spraak van sprekers die praten in een andere taal dan hun moedertaal. In het bijzonder gaat dit proefschrift over Engelse Lombard spraak door mensen wiens moedertaal Nederlands is. Ik benader dit onderzoek vanuit drie perspectieven: de akoestische eigenschappen van de Lombard spraak, de verstaanbaarheid, en hoe sterk het accent van de sprekers klinkt. Door Lombard spraak in (voor de spreker) vreemde talen te onderzoeken, heb ik meer geleerd over Lombard spraak in het algemeen, alsook over het spreken en verstaan van een vreemde taal.

Het proefschrift bestaat uit een introductie tot het onderzoeksgebied, vijf hoofdstukken die de onderzoeksvragen behandelen vanuit de drie perspectieven, en een algemene discussie. Het DELNN (Dutch English Lombard Native Non-native) corpus, dat als basis dient voor dit proefschrift, wordt beschreven in Hoofdstuk 2. Hoofdstukken 2, 3 en 4 gaan uitsluitend over de akoestische eigenschappen van Lombard spraak, evenals delen van Hoofdstukken 5 en 6. Hoofdstuk 5 gaat daarnaast nader in op de verstaanbaarheid van Lombard spraak, en Hoofdstuk 6 gaat in op de perceptie van het buitenlandse accent van de spreker.

Het DELNN Corpus

Hoofdstuk 2 beschrijft het DELNN corpus, waarvoor ik geluidsopnames heb gemaakt van 30 vrouwen die het Nederlands als moedertaal hebben, en negen vrouwen die het Amerikaans-Engels als moedertaal hebben. Deze sprekers produceerden ieder normale en Lombard spraak, waarbij de Amerikaans-Engelse sprekers dit alleen in het Engels deden, terwijl de Nederlandse sprekers dit zowel in het Nederlands als in het Engels deden. De sprekers lazen vraag-antwoord paren voor, waarbij op één van de woorden in het antwoord contrastieve focus lag. Een voorbeeld hiervan is het vraag-antwoord paar "*Bezochten de kinderen de speeltuin vanmiddag? Nee, ze bezochten de parade vanmiddag.*", waarin de contrastieve focus op het woord "parade" ligt. De locatie van deze focus was zo gekozen dat de focus ofwel op het doelwoord lag (de "late focus" conditie) ofwel op een woord voorafgaand aan het doelwoord (de "vroege focus" conditie). De doelwoorden waren zodanig gekozen dat we verwachtten dat ze moeilijk zouden zijn voor de Nederlandssprekenden wanneer ze Engels spreken. Er waren drie categorieën van doelwoorden: 1) Woorden beginnende met /θ/ (bijvoorbeeld de "th" van "think"). Deze woorden zijn lastig voor Nederlands-sprekers omdat de /θ/ niet voorkomt in het Nederlands. 2) Woorden die verwant zijn aan een Nederlands woord, waarbij de Engelse uitspraak een sjwa heeft vóór de klemtoon, en de Nederlandse uitspraak een volle

klinker heeft op die plek (bijvoorbeeld het Engelse woord “police” en het Nederlandse woord “politie”). Hierbij verwachten we dat de Nederlands-sprekers de volle klinker uitspreken in plaats van de sjwa. 3) Woorden die eindigen in een stemhebbende obstruent (bijvoorbeeld <b> of <d>). Het uitspreken van deze woorden is voor Nederlands-sprekers wellicht lastig, omdat de Nederlandse spraak “final devoicing” kent, waarbij een stemhebbende obstruent stemloos uitgesproken wordt, terwijl dit in het Engels niet gebruikelijk is (bij final devoicing wordt bijvoorbeeld een <d> uitgesproken als een /t/).

Het DELNN corpus bestaat uit de hierboven genoemde opnames, in combinatie met bijbehorende textgrids waarin de woorden en fonemen aangegeven zijn. Deze materialen werden gebruikt voor de analyse van de akoestische eigenschappen van de spraak, als stimuli voor de verstaanbaarheid-experimenten, en voor het accent-experiment. De spraakopnames en de textgrids zijn beschikbaar gemaakt voor wetenschappelijk onderzoek.

Akoestiek

Traditionele akoestische eigenschappen

Hoofdstuk 2, 3 en 4 gebruiken akoestische eigenschappen om te onderzoeken of Engelse Lombard spraak verschilt wanneer sprekers wel of niet hun moedertaal spreken, en als dit het geval is, waarom dat verschil er is. Ik heb dit aangepakt door opnames van Engelse spraak te vergelijken, tussen de mensen die het Engels als moedertaal hebben, en de mensen die het Nederlands als moedertaal hebben. Vervolgens heb ik ook nog naar de opnames gekeken van de Nederlandssprekenden en onderzocht of er verschillen waren tussen hoe Lombard spraak geproduceerd werd in het Nederlands en in het Engels. In Hoofdstuk 2 heb ik onderzoek gedaan naar de volgende eigenschappen: intensiteit, spectral Center of Gravity, woord-duur en voice onset time (VOT). Het resultaat hiervan was dat deze vier eigenschappen op een gelijke manier veranderen voor beide groepen sprekers in Engelse Lombard spraak, in vergelijking met normale spraak. Echter, de groep met het Nederlands als moedertaal verschilde wel in intensity, spectral Center of Gravity en VOT, wanneer deze groep Engels sprak versus Nederlands. Wanneer de groep met Nederlands als moedertaal Lombard spraak produceert in het Nederlands, laten ze een grotere verhoging in spectral Center of Gravity zien, een kleinere verhoging in intensiteit voor de late focus conditie, en een kleinere verlaging in VOT, in vergelijking met wanneer deze groep Engels spreekt.

Hoofdstuk 3 onderzocht de mediaan van de grondfrequentie (F0). Hieruit blijkt dat de sprekers met het Nederlands als moedertaal hun F0 verhogen in Lombard spraak, in zowel de vroege als de late focus condities, in zowel het Nederlands als in het Engels. De grootte van dit effect is vergelijkbaar in beide talen. In het geval van late focus verhoogden de twee groepen sprekers van het Engels hun F0 in vergelijkbare mate in Lombard spraak, maar in

geval van vroege focus was er een verschil. In geval van vroege focus lieten de sprekers met het Engels als moedertaal geen significante verhoging in F0 zien, terwijl dit voor de sprekers met Nederlands als moedertaal wel het geval was.

Hoofdstuk 4 onderzocht de variatie in het bereik van de grondfrequentie (F0), en concludeerde dat het bereik afhankelijk is van de locatie van focus. Voor de late focusconditie verhoogde het frequentiebereik het meeste voor Engelse spraak geproduceerd door de groep met het Engels als moedertaal, gevolgd door de Engelse spraak geproduceerd door de groep met het Nederlands als moedertaal, en tot slot de Nederlandse spraak.

Ik beschouw de resultaten van zowel Hoofdstuk 3 als 4 als een invloed van de moedertaal op de productie van Lombard spraak in een vreemde taal. Deze resultaten geven de indruk dat ook de sprekers voor wie Engels niet de moedertaal was, toch Lombard spraak produceren in het Engels. Bovendien lijken er subtiele verschillen te zijn in de Lombard spraak van deze sprekers in vergelijking met de groep die Engels wel als moedertaal heeft, op het gebied van enkele akoestische eigenschappen.

Computationele akoestiek

Delen van hoofdstuk 5 en 6 bepaalden de akoestische eigenschappen met behulp van computationele maten. Hoofdstuk 5 gebruikte de "High Energy Glimpsing Proportion" (HEGP; Tang & Cooke, 2016) metriek om een selectie van woorden uit het DELNN corpus te analyseren, die uitgesproken zijn door acht sprekers wiens moedertaal Engels is, en acht sprekers voor wie dat niet het geval is. De HEGP metriek geeft aan hoe goed de spraak bestand is tegen achtergrondgeluid. Hoofdstuk 5 laat zien dat Lombard spraak een hogere HEGP heeft dan normale spraak, voor beide sprekersgroepen.

Hoofdstuk 6 maakte gebruik van een forced aligner om nader te kijken naar bepaalde fonemen in de drie voorgenoemde categorieën van de Engelse doelwoorden. De forced aligner kon voor de spraakopnames kiezen uit de prototypische Engelse uitspraak, of een Nederlandse variant. De doelwoorden werden geselecteerd uit twee condities in het DELNN corpus, die maximaal van elkaar verschilden in "articulatory effort". Voor één conditie werd het doelwoord uitgesproken in Lombard spraak met contrastieve focus. Deze conditie werd GE genoemd, voor "great effort" (veel moeite). In de andere conditie werd het doelwoord uitgesproken in normale (niet-Lombard) spraak zonder contrastieve focus. Deze conditie werd LE genoemd, voor "less effort" (minder moeite). Over het algemeen kan ik concluderen dat de sprekers met het Nederlands als moedertaal minder fonemen op de prototypisch Engelse manier uitspreken, volgens de forced aligner. De twee sprekersgroepen gedragen zich vergelijkbaar tussen de LE en GE condities. De forced aligner vond alleen een verschil in de "sjwa" doelwoordcategorie, waarin er minder prototypische "sjwa" fonemen gevonden

werden in GE dan in LE, voor beide sprekersgroepen.

Verstaanbaarheid

Het doel van Hoofdstuk 5 was om te achterhalen of Lombard spraak, geproduceerd door iemand die een vreemde taal spreekt, beter verstaanbaar is dan normale spraak, wanneer deze spraak gehoord wordt in een rumoerige omgeving. Als dit het geval was, was het volgende doel om te achterhalen of, en op welke manier, de combinatie van de moedertalen van de sprekers en de luisteraars hier invloed op heeft. Hiervoor zette ik een experiment op met drie groepen luisteraars die verschillen in hun moedertaal (Nederlands, Spaans of Engels). Deze groepen luisterden naar opnames van Engelse woorden, die gepresenteerd werden in ruis. De opnames verschilden in twee dimensies: ze waren uitgesproken door moedertaalsprekers van het Nederlands of het Engels, en ofwel als Lombard ofwel als normale spraak. Ongeacht de moedertaal van de spreker, was Lombard spraak beter te verstaan dan normale spraak. Opmerkelijk genoeg vond ik geen verschillen in de grootte van dit effect tussen de sprekers die het Engels en die het Nederlands als moedertaal hebben. Dit suggereert dat de aanpassingen die de sprekers maken bij Lombard spraak voordelig zijn voor de luisteraars, ongeacht de moedertaal van de spreker. Tot slot bleek dat de luisteraars met een andere moedertaal dan het Engels het meeste baat hebben bij Lombard spraak, maar binnen deze groep vond ik geen verschil tussen de Nederlandse en Spaanse luisteraars.

Accent

Hoofdstuk 6 onderzocht hoe het accent van de spreker waargenomen wordt in de GE en LE condities die, zoals hierboven beschreven, verschillen in articulatory effort. Hiervoor werd een luisterexperiment opgezet waarin twee versies van een aantal Engelse doelwoorden afgespeeld werden, ingesproken door dezelfde spreker. De ene versie van een doelwoord was uitgesproken in de GE conditie en de ander in de LE conditie. De opnames die gebruikt werden voor dit experiment zijn dezelfde als die ook gebruikt werden voor het onderzoek met de forced aligner, en bestaan dus uit de drie categorieën van doelwoorden: woorden die beginnen met /θ/, woorden die verwant zijn aan een Nederlands woord maar die een sjwa hebben voor de klemtoon, en woorden die eindigen op een stemhebbende obstruent. Net als voorheen zijn een deel van deze opnames uitgesproken door sprekers met het Engels als moedertaal, en een deel door sprekers met het Nederlands als moedertaal. Voor dit experiment waren er twee groepen luisteraars: de ene groep had het Engels als moedertaal, de andere het Spaans. De taak van de luisteraar was om aan te geven welke van de twee versies van een doelwoord, uitgesproken door dezelfde spreker, een minder sterk accent had: de LE- of de GE-variant. Voor de groep luisteraars die het Engels niet als moedertaal had (en dus Spaans), vond ik geen

onderscheid tussen de GE of LE condities wat betreft accent, voor beide sprekersgroepen. Voor de groep luisteraars die wel het Engels als moedertaal had, was dit anders. Sprekers die het Engels als moedertaal hadden, vonden de LE conditie minder geaccentueerd voor de sjwa en de stemhebbende woordfinale obstruent categorieën, terwijl ze geen significante voorkeur voor GE of LE hadden in de /θ/ categorie. Voor sprekers die het Nederlands als moedertaal hebben en luisteraars die het Engels als moedertaal hebben, gold deze voorkeur voor LE wel voor de stemhebbende woordfinale obstruent categorie, maar niet voor de andere twee woordcategorieën. In het algemeen beoordeelden beide luisteraarsgroepen de sprekersgroepen vergelijkbaar, en was er alleen een verschil tussen de twee condities voor de doelwoorden in de sjwa categorie. Dat de moedertaal van de sprekers nauwelijks invloed had op de voorkeur voor LE of GE voor beide groepen luisteraars, suggereert dat de sprekers die het Engels niet als moedertaal hebben vergelijkbare akoestische veranderingen maken als sprekers die het Engels wel als moedertaal hebben, wanneer ze Lombard spraak produceren.

Conclusie

Dit proefschrift documenteert de eigenschappen van Engelse Lombard-spraak uitgesproken door moedertaalsprekers van het Nederlands. Hoewel er een subtiel verschil is tussen de aanpassingen in Engelse Lombard spraak tussen deze sprekers en sprekers die het Engels wel als moedertaal hebben, zijn de aanpassingen zeer vergelijkbaar.

Acknowledgements

Thank you Mirjam and Esther for taking a chance on me even though I mixed up the time zones and was late for my PhD interview. As cliché as it sounds, this project wouldn't have been possible without both of you. Mirjam, having a meeting with you always left me excited about the results, especially those involving schwa. Thank you for helping me gain perspective on the project and being so supportive. You are a great mentor, and I always found your advice insightful. Apart from your academic input, I appreciated your efforts to integrate me into the Dutch culture by explaining Dutch traditions in my first few years and later starting our meetings with some koetjes en kalfjes in Dutch. Not to mention, your on-point museum recommendations. Esther, thank you for finding all my weaknesses and pointing them out in the kindest possible way. Despite only meeting once a month, I learned a great deal from you and wanted to thank you for always being available and willing to help. The two of you together were a wonderful supervision team, thanks again for making this thesis possible.

I wanted to thank the manuscript committee, Marc, Miquel, Mirjam, Rob and Valerie for taking the time to carefully read through my thesis. Not to mention, for responding to my many emails promptly, allowing this defense to come together before the summer.

Many people have made the PhD experience a much smoother and more enjoyable ride. Margret, thank you for your organizational skills and being the backbone of the CLS lab. You made running participants a smooth and painless process. Kevin, thank you for organizing courses that were perfectly tailored to the PhD experience, they were essential to my research. Furthermore, thanks for all the IMPRS social events with New York pizza, it was great to see everyone outside of work and it was a nice way to meet new people. Martin and Ben thank you for hosting me in Vitoria and Edmonton and for co-authoring our article. It was my first full article from my PhD project and it was great to work with you.

Many thanks to the Speech Production and Comprehension group – Amalia, Annika, Aurora, Chen, Cong, Emily, Esther, Hanno, Joe, Katie, Lieke, Lisa, Lotte, Louis, Martijn, Mirjam, Robert, Stella, Sophie, Tim, Yachan, and Xinyu - for being a welcoming group to share ideas with and receive critical feedback from. Louis, thank you for always being willing answer my many questions. I learned so much from you about all the technical aspects of my research and really appreciate that you always made time to sit down and talk. Elisabeth, it was fun to work with you on your master's thesis and I hope that you also enjoyed the experience and learned a lot. Thank you to the SPRINT team, Amalia, Cong, Katie, and Stella, for your encouragement and support in the last year of my PhD.

A special thanks goes to my paranymphs, Aurora and Lotte. It has been wonderful having you two by my side these past five years, with random conversations during our lunches, much needed fancy coffee breaks, game nights, paddle games, and fun days out.

Aurora, thanks for keeping me young and up to date with all the cool new lingo and trends, TikTok, how to RT and always having the perfect playlist for any occasion. Lotte, thanks for your cheerful spirit, boundless energy, and exquisite taste in food. Furthermore, thanks for being the best driver in the Netherlands and Tasmania (which was way larger than estimated and hence we had much more driving than expected).

My whole PhD wouldn't be the same without Chen, who has been there every step of the way. I am very thankful to have you as a friend and as an office-mate. I had a blast traveling to all our ENRICH events and other conferences together. I'm so happy to have started and be ending this journey together.

Annika and Lisa, thanks for also being wonderful officemates. Thank you for showing me the ropes when I arrived and for all the lovely coffee breaks and lunches. Of course, coffee breaks and lunches wouldn't be the same without all the other 8th and 9th floor PhDs – Aurélia, Chantal, Claire, Elly, Ferdy, Figen, Gert-Jan, Hannah, Hanno, Imke, Maria, Patricia, Saskia, Tashi, Theresa, Thijs, Tim, Wei, Xiaoru and Yu.

ENRICHies, thanks so much for all the fun adventures we had during out network meetings all over Europe. Spending three weeks in a villa in Crete was pretty awesome. It was fun learning with and from you. I'm so excited that now four of us are living in the Netherlands and we can meet up.

Emily, thanks for your friendship, book recommendations, healthy meals (which has inspired me to try whole-wheat pasta), and good conversations. It's nice to bond over how parallel our lives are: Americans moving to the Netherlands, doing a PhD in linguistics, joining the same book club, finding a Dutch partner and buying a house in Nijmegen – it's funny how you are always one step ahead, good to know what's coming.

Candice, so nice to be living in the same city again, and reunited after our Master's in Barcelona. We may have aged a few years since our Master's, but I'm sure we will manage to make it out and party a few times. I'm looking forward to having you around and hanging out together.

Lia and Marta, I am so glad that I decided to go to the market one Saturday morning and met you two! Spending time with both of you, during the first part of my PhD was so much fun and a great introduction to the Netherlands. Thanks for all the fun memories – King's Day with karaoke on stage, our trip to Belgium, and our final "dinner" together when I failed spectacularly to make mac-n-cheese so we instead had cheesecake at Hotel Credible. Looking forward to visiting both of you in Spain!

Anne and Mary, you two are the best roommates I could ask for – it truly felt like we were a family living together – having dinners, celebrating Sinterklaas, birthdays, and going on walks together. I am so glad that we always made an effort to meet even after I moved out.

Spinning class on Friday with you was the best way to start the weekend.

I am grateful for all the activities and the people involved in them that kept me sane during my PhD – most importantly acro. Thank you to everyone who I have jammed with for helping me be in the moment and for throwing me as high as I could go. Acro classes are always filled with so much laughter and fun that I come away feeling better. Thanks as well to all my friends outside acro for listening to my acro stories and watching my solo re-enactments without laughing too much. In the past couple years, I also joined Book club. Sarah, thank you for organizing it, and for everyone for such lovely conversations that may or may not involve the book.

Beat Owls, although we are physically far apart, you are close to my heart. Thank you Betrice, Darcey, Julie, Sarah and Subin for all the five-o'clock dinners which I dragged you to in my hangry state, zoom calls across too many time zones, and laughter. Thanks for all the memories, even the one of that New Year's Eve when we decided to go to Times Square where we stood in line for hours with bursting bladders, freezing in the cold and overpaying for a pizza. It means a lot to have life-long friends who are there for each other through the good and the bad.

Thank you to Andre, Annette, Tineke, Lennart, Jasper, Felix, Cecile, Leon, and Marieke for welcoming me into the van Lümig family. I knew I had great parents-in-law when you moved my belongings into our new apartment when I was away on a work trip and when you did klussen on our house when we were away on vacation.

I want to thank my parents, who have always been very supportive, and have never once complained about the fact that their only child lives so far away from them. Also, thank you for sparking my love for reading. I remember that my dad would read Harry Potter to my mom and I while she cooked. I soon learned that it would go faster if I read it alone and I never stopped reading after that. Mom, I love enjoying the outdoors with you and being active together. I still fondly remember that one summer that you managed to convince me to spontaneously cycle 100 miles (161 km). Dad, thank you for always offering to read through my work to help me improve – from high school essays, to job applications, you and mom are always willing to help whenever and however you can. Dad, seeing you finish your PhD while working full time was an inspiration, which I have an even greater appreciation of now that I am finishing mine. I hope that both of us being Dr. Marcoux won't be too confusing.

Arno, thank you for being by my side rain or shine these past four and a half years – and considering that we live in the Netherlands, it was mostly rain. Your support with my thesis, especially in the final stages has been a huge help. If we managed to buy a house and make it our home while I was in the last stages of my PhD, nothing can stop us. You are my favorite person to pair program with, not to mention my favorite person in general.

Curriculum Vitae

Katherine Marcoux was born in Rochester, New York in the United States in 1992. In 2014 she completed her Bachelor's (summa cum laude) with honors in her in Linguistics and Languages (Japanese and Spanish) major and a minor in Psychology from Bryn Mawr College. She spent one year in Andorra on a Fulbright Teaching Assistantship. Following this, she moved to Barcelona and obtained her Masters of Science from the University Pompeu Fabra. In 2017 she began her PhD as an early-stage researcher in the European Training Network ENRICH, funded by EU's H2020 research and innovation programme. She conducted her research at the Centre for Language Studies of the Radboud University, where she was a part of the Graduate School of Humanities and the International Max Planck Research School for language Sciences. Last year she began her current position as lab manager in the SPRINT team at Radboud University.

Publications

PhD related

Marcoux, K., Cooke, M., Tucker, B. V., & Ernestus, M. (2022). The Lombard intelligibility benefit of native and non-native speech for native and non-native listeners. *Speech Communication*, *136*, 53-62. https://doi.org/10.1016/j.specom.2021.11.007

Marcoux, K., & Ernestus, M. (2019a). Differences between native and non-native Lombard speech in terms of pitch range. In M. Ochmann, M. Vorländer, & J. Fels (Eds.), *Proceedings of the ICA 2019 and EAA Euroregio. 23rd International Congress on Acoustics, integrating 4th EAA Euroregio 2019* (pp. 5713–5720). Berlin, Germany: Deutsche Gesellschaft für Akustik. https://doi.org/10.18154/RWTH-CONV-239240

Marcoux, K., & Ernestus, M. (2019b). Pitch in native and non-native Lombard speech. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2605–2609). Melbourne, Australia: Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Marcoux, K., & Ernestus, M. (submitted). Acoustic characteristics of non-native Lombard speech in the DELNN corpus.

Marcoux, K., Süß, E., & Ernestus, M. (in prep). How articulatory effort affects the strength of non-native speakers' accentedness.

Other

Arvaniti, A., Gryllia, S., Zhang, C., & Marcoux, K. (2022). Disentangling emphasis from pragmatic contrastivity in the English H* ~L+H* contrast. In *Proceedings of Speech Prosody 2022.*

Baus, C., McAleer, P., Marcoux, K., Belin, P., & Costa, A. (2019). Forming social impressions from voices in native and foreign languages. *Scientific reports*, *9(1)*, 1-14.

Gryllia, S., Arvaniti, A., Zhang, C., & Marcoux, K. (2022). The many shapes of H*. In *Proceedings of Speech Prosody 2022.*