PITCH IN NATIVE AND NON-NATIVE LOMBARD SPEECH

Katherine Marcoux and Mirjam Ernestus

Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
k.marcoux@let.ru.nl; m.ernestus@let.ru.nl

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.

Keywords: Lombard speech, native speech, non-native speech, pitch, English

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., 5, 13, 17, 23]. Research done to date has extensively studied native Lombard speech, primarily in English [e.g., 17, 23], but also in other languages including Spanish [e.g., 5] and French [e.g., 11]. 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., 2, 8, 10]. 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., 15, 21]. 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 [24]). This mean pitch is higher for American-English (AmE) women of a similar age at 224 Hz [22]. Research on bilinguals further illustrates mean pitch differences among languages. For instance, Voigt, Jurafsky, and Sumner [25] examined German-French and German-Italian bilinguals, finding that individuals had different mean pitches in their two languages. Finally, several studies [e.g., 20] 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., 12]. Importantly, RP and Dutch speakers have similar pitch accents at the sentence level in their native languages [e.g., 12]. 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., 26]. Post-focus compression narrows and lowers the pitch range for material after the focused word.

2. METHODS

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 [7], as indicated by their LexTALE [16] 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.

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.

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 [4]) 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 [16] 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.

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 [3]. 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., 14]. 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).

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 [1] in R [19]. 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.

2.6 Results

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.

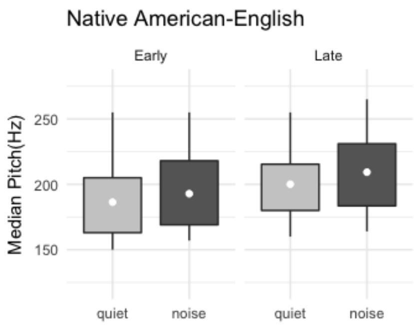
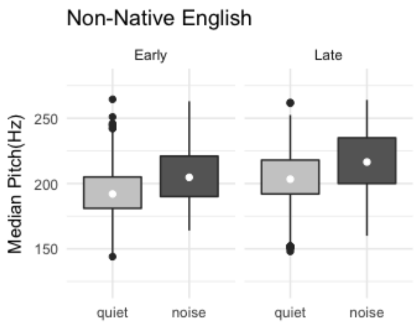
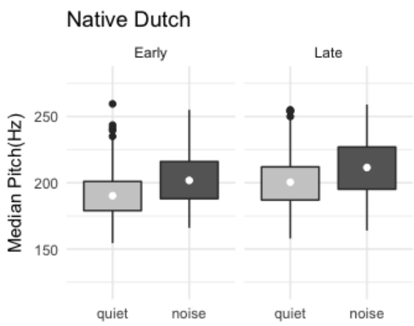


Figure 2: Boxplot of the median pitch values of non-native English in early- and late-focus in quiet and noise. The white dots represent the means.



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. 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., 23].

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., 22] than native Dutch pitch [e.g. 24].

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., 26].

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., 9, 17, 18, 23]. This Lombard benefit is smaller for non-native listeners [e.g., 6]. 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.

4. ACKOWLEDGEMENTS

This project was funded by the European Union's Horizon 2020 research innovation programme (Marie Skłodowska-Curie grant No. 675324). Special thanks to Esther Janse for all her help with feedback and comments.

5. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. of Statistical Software*, 67(1), 1-48.

[2] Best, C. T. 1995. A Direct Realist Perspective on Cross-Language Speech Perception. In: Strange, W. (ed), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language research*, 167-200.
[3] Boersma, P. and Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved from http://www.praat.org
[4] Brüel & Kjær Artificial Ear/Ear Simulator (IEC 60318-1 Coupler) https://www.bksv.com/en/products/transducers/ear-simulators/ear-mouth-simulators/4153
[5] Castellanos, A., Benedí, J.M., Casacuberta, F. 1996. An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Com*. 20(1–2), 23–35.
[6] Cooke, M., Lecumberri, M. L. G. 2012. The intelligibility of Lombard speech for non-native listeners. *J. Acoust. Soc. Am*. 132(2), 1120–1129.
[7] Council of Europe. 2011. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Retrieved from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp
[8] Cutler, A. 2012. Second-Language Listening: Sounds to Words. In*: Native listening: Language experience and the recognition of spoken words*. London: MIT Press, 303-335.
[9] Dreher, J. J., O'Neill, J. 1957. Effects of Ambient Noise on Speaker Intelligibility for Words and Phrases. *J. Acoust. Soc. Am*. 29(12), 1320–1323.
[10] Flege, J. E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *J. Phon.* 15(1), 47-65.
[11] Garnier, M., Henrich, N. 2014. Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Comp. Speech & Lang*. 28(2), 580–597.
[12] Gussenhoven, C., Broeders, A. 1997. Intonation. In: *English pronunciation for student teachers.* Groningen: Noordhoff Uitgevers, 174-188.
[13] Junqua, J.C. 1993. The Lombard reflex and its role on human listeners and automatic speech recognizers. *J. Acoust. Soc. Am*. 93(1), 510–524.
[14] Keating, P., Garellek, M., Kreiman, J. 2015. Acoustic properties of different kinds of creaky voice. *Proc. 18th ICPhS Glasgow*, 0821-1.
[15] Kormos, J. 2006. Monitoring. In: *Speech Production and Second Language Acquisition*. London: Lawrence Erlbaum Associates, Publishers, 122-136.
[16] 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.
[17] Lu, Y., Cooke, M. 2008. Speech production modifications produced by competing talkers, babble, and stationary noise. *J. Acoust. Soc. Am*. 124(5), 3261–3275.
[18] Pittman, A. L., Wiley, T. L. 2001. Recognition of Speech Produced in Noise. *J. of Speech, Lang., and Hearing Research*, 44(3), 487–496.
[19] R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
[20] Rasier, L., Hiligsmann, P. 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux cahiers de linguistique française*, 28, 41-66.
[21] Segalowitz, N. 2010. Second Language Cognitive Fluency. In: *Cognitive bases of second language fluency*. London: Routledge, 74-106.
[22] Stoicheff, M. L. 1981. Speaking fundamental frequency characteristics of nonsmoking female adults. *J. of Speech, Lang., and Hearing Research*, 24(3), 437-441.
[23] Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., Stokes, M. A.. 1988. Effects of noise on speech production: Acoustic and perceptual analyses. *J. Acoust. Soc. Am*. 84(3), 917–928.
[24] Van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Lang. and speech*, 38(3), 253-265.
[25] Voigt, R., Jurafsky, D., Sumner, M. 2016. Between-and Within-Speaker Effects of Bilingualism on F0 Variation. *Proc. Interspeech San Francisco,* 1122-1126.
[26] Xu, Y. 2011. Post-focus compression: Cross-linguistic distribution and historical origin. *Proc. 17th ICPhS Hong Kong,* 152-155 .