PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

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

Katherine MARCOUX[1]; Mirjam ERNESTUS[2]

Radboud University, the Netherlands

ABSTRACT

Lombard speech, speech produced in noise, is acoustically different from speech produced in quiet (plain speech) in several ways, including having a higher and 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.

Keywords: Lombard speech, non-native, pitch range

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., 1, 2, 3, 4). Some researchers have also found that Lombard speech is accompanied by a larger pitch range (e.g. 5, 6). The extensive research on Lombard speech to date has focused on native Lombard speech in various languages, including English (e.g. 3, 4), French (e.g. 5), and Spanish (e.g. 1). 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., 7, 8, 9), and prosody (e.g.,10). 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. 11, 12). 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 compared to native speakers. This is in line with an earlier study (13), 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,

[1] k.marcoux@let.ru.nl
[2] m.ernestus@let.ru.nl

which is in line with what the Dutch do in their native Dutch (13).

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 (14) 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 (15), while similarly aged Dutch females were found to have a lower mean pitch of 191 Hz (as cited in 16). These differences in mean pitch per language are unlikely to be due to physical differences, as Voigt, Jurafsky, and Sumner (17) 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.

2. METHODS

2.1 Participants

Marcoux and Ernestus' (13) Dutch English Lombard Native Non-Native (DELNN) corpus was used for the present study. 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 (18), as indicated by their LexTale (19) 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.

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.

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 (20). 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 (19) 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.

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 (21) 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 (22). 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 (13).

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 (23). For the conversion, we calculated semitones with the following equation, as was used by Kitamura and colleagues (24):

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

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 (25) in R (26), 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 ANOVA on nested models and AIC scores on non-nested models. In the case of ANOVA, 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 (27) 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.

3. RESULTS

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 of the sentences. 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.

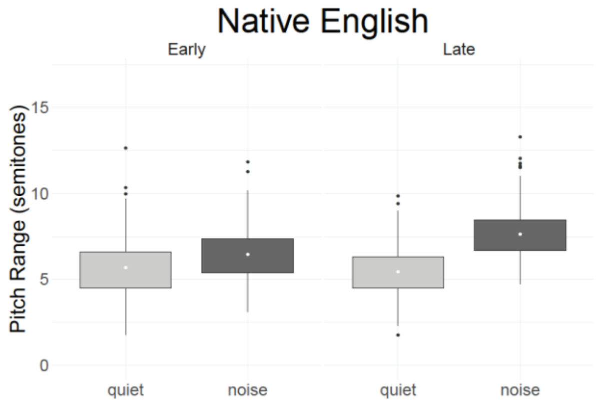

Figure 1 – A boxplot of pitch range of native English data for early- and late-focus for quiet and noise conditions. The means are represented by the white dots.

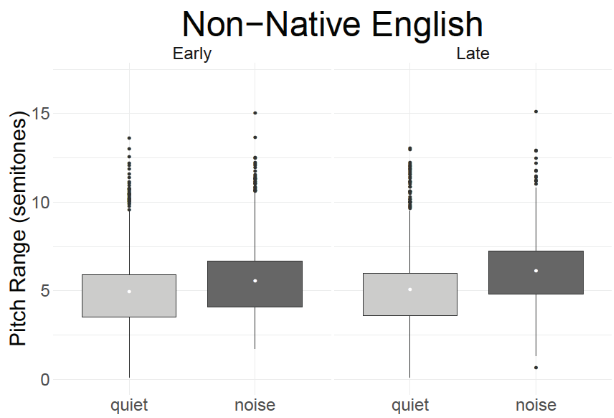

Figure 2 – A boxplot of pitch range of non-native English data for early- and late-focus for quiet and noise conditions. The means are represented by the white dots.

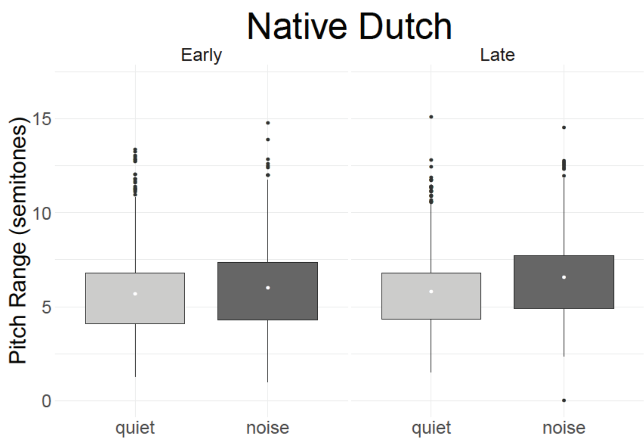

Figure 3 – A boxplot of pitch range of native Dutch data for early- and late-focus for quiet and noise conditions. The means are represented by the white dots.

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		-		-
Residual		0.98		1.06		0.88

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 full model (Table 1).

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 native 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
Residual		1.06

4. DISCUSSSION

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., 5, 6), 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 (13), 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 (28). 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 (14). 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.

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 (13). 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.

ACKNOWLEDGEMENTS

This project has received funding from the European Union’s Horizon 2020 research innovation programme under the Marie Skłodowska-Curie grant agreement No. 675324. Special thanks additionally go to Esther Janse for her help in shaping the project, with her insight, comments, and feedback.

REFERENCES

1. Castellanos A, Benedí JM, Casacuberta F. An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. Speech Com. 1996;20(1-2):23-35.
2. Junqua JC. The Lombard reflex and its role on human listeners and automatic speech recognizers. The J Acoust Soc Am. 1993; 93(1):510-24.
3. Lu Y, Cooke M. Speech production modifications produced by competing talkers, babble, and stationary noise. J Acoust Soc Am. 2008; 124(5):3261-75.
4. Summers WV, Pisoni DB, Bernacki RH, Pedlow RI, Stokes MA. Effects of noise on speech production: Acoustic and perceptual analyses. J Acoust Soc Am. 1988; 84(3):917-28.
5. Garnier M, Henrich N. Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?. Computer Speech & Language. 2014;28(2):580-97.
6. Welby P. Intonational differences in Lombard speech: Looking beyond F0 range. Proc Third International Conference on Speech Prosody; 2 May 2006; Dresden, Germany 2006. p. 763-766.
7. Best CT. A Direct Realist View of Cross-Language Speech Perception. In: Speech perception and linguistic experience: Theoretical and methodological issues in cross-language research. 1995:171-204.
8. Cutler A. Second-Language Listening: Sounds to Words. In: Native listening: Language experience and the recognition of spoken words. London, UK: MIT Press; 2012. p. 303-335.
9. Fledge JE. The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. J Phon. 1987;15(1):47-65.
10. van Maastricht L, Krahmer E, Swerts M. Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa. Lang Learning. 2016;66(1):124-58.
11. Kormos J. Monitoring. In: Speech Production and Second Language Acquisition. London, UK: Lawrence Erlbaum Associates, Publishers; 2006. p. 122-136.
12. Segalowitz N. Second Language Cognitive Fluency. In: Cognitive Bases of Second Language Fluency. London, UK: Routledge; 2010. p.74-106.
13. Marcoux K, Ernestus M. Pitch in native and non-native Lombard speech. Proc ICPhS; 5-9 August 2019; Melbourne, Australia 2019.
14. Gussenhoven CH, Broeders A. Intonation. In: English pronunciation for student teachers. 2nd ed.

Groningen/Houten, the Netherlands: Noordhoff Uitgevers; 1997. p.174-188.
15. Stoicheff ML. Speaking fundamental frequency characteristics of nonsmoking female adults. J Speech, Lang, and Hearing Research. 1981;24(3):437-41.
16. van Bezooijen R. Sociocultural aspects of pitch differences between Japanese and Dutch women. Lang and speech. 1995;38(3):253-65.
17. Voigt R, Jurafsky D, Sumner M. Between-and Within-Speaker Effects of Bilingualism on F0 Variation. Proc Interspeech; 8-12 September 2016; San Francisco, USA 2016. p. 1122-1126.
18. Council of Europe. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. 2011. Retrieved from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp
19. Lemhöfer K, Broersma M. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. Behavior research methods. 2012;44(2):325-43
20. Brüel & Kjær Artificial Ear/Ear Simulator (IEC 60318-1 Coupler) https://www.bksv.com/en/products/transducers /ear-simulators/ear-mouth-simulators/4153
21. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.0.37. 2018. Retrieved from http://www.praat.org
22. Keating PA, Garellek M, Kreiman J. Acoustic properties of different kinds of creaky voice. Proc ICPhS; 10-14 August 2015; Glasgow, United Kingdom 2015.
23. Mennen I, Schaeffler F, Docherty G. Pitching it differently: A comparison of the pitch ranges of German and English speakers. Proc ICPhS; 6-10 August 2007; Saarbrücken, Germany 2007.
24. Kitamura C, Thanavishuth C, Burnham D, Luksaneeyanawin S. 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. 2001; 24(4):372-92.
25. Bates D, Maechler, M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J. of Statistical Software. 2015;67(1):1-48.
26. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2016; Vienna, Austria.
27. Baayen RH. Analyzing linguistic data: A practical introduction to statistics using R. Cambridge, UK: Cambridge University Press; 2008.p. 165-240
28. Xu Y, Xu CX. Phonetic realization of focus in English declarative intonation. J Phon. 2005;33(2):159-97.