INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Prosodic Characteristics of Mandarin Declarative and Interrogative Utterances in Parkinson's Disease

Lei Liu[1], *Meng Jian*[1], *and Wentao Gu*[1,2]

[1]School of Chinese Language and Culture, Nanjing Normal University, China
[2]Nanjing Normal University and the George Washington University Joint Laboratory of Speech, Hearing and Rehabilitation Sciences

1990liulei@163.com, jennifer7fish@gmail.com, wtgu@njnu.edu.cn

Abstract

This work investigated the prosodic characteristics of declarative and interrogative utterances produced by speakers with Parkinson's disease (PD), in comparison to healthy controls (HC). Forty native speakers of Mandarin, including 20 PDs and 20 age-matched HCs, recorded 32 utterances varying in sentence type, sentence length, and sentence-final tone. SS-ANOVA was used to show the F0 contours and the global and final-syllable F0 level, F0 slope, speech rate, and intensity ratio were statistically analyzed using linear mixed-effects models. For the HC group, interrogative utterances showed a significantly higher mean F0 than declarative utterances. The PD group exhibited no significant F0 difference between declarative and interrogative utterances, coinciding with our subjective impression on PD's monotonous voice of tone. This suggests that PD's ability to control fundamental frequency degraded in comparison to HC. Also, the PD group produced significantly faster speech, especially final syllable, than the HC group, suggesting that PD's articulatory control degraded at the end of an utterance.

Index Terms: Parkinson's disease, Mandarin speech, declarative, interrogative, prosody

1. Introduction

Parkinson's disease (hereafter PD) is one of the neurodegenerative diseases in the middle-aged and elderly people, with typical motor impairments such as bradykinesia, hypokinesia, akinesia, muscle rigidity, and rest tremor [1]. In addition, 70-90% of PD patients also suffer from hypokinetic dysarthria [2], which may have developed for years before the appearance of obvious clinical motor symptoms and hence may be an indicator for early diagnose of PD. The hypokinetic dysarthria in PD is manifested in all dimensions of speech production, especially in speech prosody [3].

Prosody plays important roles in speech communication. Statements and unmarked yes-no questions (intonation question) are associated with falling and rising sentence intonation, respectively. Their patterns of F0 raising in questions, however, are different. Mandarin raises F0 in a global domain (viz., questions start at a higher F0 than statements) [4, 5] – though F0 rising may still peak at the end of the sentence [6]. Mandarin syllables in questions have shorter duration than those in statements, except at the final position where syllable duration is longer in questions [5, 7]. Questions have higher intensity than statements, and the difference between them increases until the end of the sentences [5].

The results of previous studies about PDs' prosodic parameters are not consistent. As for F0, some studies showed that PD speakers had significantly higher average F0 than the healthy controls (hereafter HC) [8-10]. However, [11, 12] did not observe F0 mean difference between the two groups. Studies on overall speech rate of the PD group remained inconclusive, which was abnormally fast [13], abnormally slow [14] or same as control speakers [8, 12, 15]. In regard to intensity, reduced [9, 16] or equivalent [8] intensity were found. Inconsistent conclusions above not only show the large individual difference of PD patients but also reflects the complexity of PD.

It is to be noted that although the German-speaking PD produced more monotonous prosody, they showed similar intonation contrasts between statement and question, as in HC speakers [10]. Likewise, Cantonese-speaking PD used similar acoustic cues as HC speakers to mark the question–statement contrast, though the contrast was not observed in all speakers [17]. German is a non-tone language, and thus there is no interaction between tone and intonation. For Cantonese, listeners mainly used F0 cues at the final syllable for intonation identification [17, 18], it may be sufficient for Cantonese speakers with PD to handle the final syllable well to ensure the intonation contrasts. While the language-specific global rising of Mandarin question [4, 5] poses longer domain F0 handling ability for Mandarin PD speakers.

The purpose of this study was to investigate intonation contrast in Mandarin speakers with PD. Different sentence lengths and final-syllable tones were inspected systematically, which reflected the dynamic change of intonation.

2. Method

Two sets of Twenty native Mandarin speakers with idiopathic PD (10M and 10F) were recruited from Nanjing Brain Hospital. They had the mean age of 66.75 years. Before speech recording, all patients were assessed by neurologists using a series of scales, including the third part (UP III) and the fifth part (i.e., modified H&Y) of the Unified Parkinson's Disease Rating Scale (UPDRS), Mini-Mental State Examination (MMSE). None of them had dementia or a reported history of other neurological diseases. Twenty healthy native Mandarin speakers served as the controls, matching the PD group on the age, gender, place of origin, and education level. Healthy control aged between 58 and 78 years old, with an average age of 64.4 years. Student's t-test shows no statistically significant difference between two groups of the age. Table 1 summarizes the basic information of the speakers.

http://dx.doi.org/10.21437/Interspeech.2019-3276

Table 1: Basic information of the PD and HC groups (Mean ± SD).

	PD	HC
Age (year)	66.75±6.00	63.4±4.76
Modified Hoehr & Yahr	2.85±0.84	n/a
MMSE	26.2±2.71	n/a
UPDRS III	30.45±8.51	n/a
UPDRS III-18	1.1±1.08	n/a

A set of Mandarin sentences varying in sentence type (statement vs. question), sentence length (3, 5, 7, or 9 syllables) and sentence-final tone (T1, T2, T3, or T4) were designed. All non-final syllables were in T1. Each target sentence was embedded in a dialogue, for which a prompt text was provided to elucidate the scenario and the relationship between the two interlocutors. Each dialogue consisted of 3 to 6 turns, and the target sentence constituted a single turn by itself to ensure that the target sentence has a wide focus.

Speech data collection was conducted in a quiet room, using a portable digital recorder Zoom H5, digitized at 44.1 kHz and 16-bit precision. The experimenter played Role A, while the participants played role B. The distance between the microphone and the test lips is about 15 cm. During recording, all PD speakers are at the medication-off state. After recording, a total of 1280 target utterances (i.e., 40 participants * 2 sentence types * 4 sentence lengths * 4 final tones) were extracted from the recorded speeech for data analysis.

3. Data analysis

Segmentation of the recorded utterances into syllable initials and rhymes was done manually. F0 values were extracted at 10ms intervals using autocorrelation analysis in Praat [19]. After manual correction of gross errors in F0 extraction, the raw F0 values were smoothed and interpolated within syllable rhymes. Ignoring durational differences, the rhyme-based time-normalized F0 contours were obtained by extracting the F0 values at 10 equidistant points in the rhyme of each syllable. The F0 values were measured in the scale of semitone (St) by setting the reference at 50 Hz. As for F0 slopes, the simple linear regression of F0 as a mathematical function of the corresponding time point was modeled for each target utterance, and the slope was obtained from the regression. The duration of each syllable was also extracted by Praat. The mean intensity value of each syllable (in dB SPL) was measured from the 20% to the 80% of the voiced segment by an averaging method using the algorithm in the Praat.

3.1. SS-ANOVA for F0 contour

A Smoothing Spline ANOVA (SS-ANOVA) [20] was conducted on the F0 data to generate the contours which show the smoothing spline fit for sentences in different groups, sentence types, sentence lengths, and final-syllable tones. SS-ANOVA is a statistical technique developed specifically for the comparison of contours along with multiple reference points, which has been proved to be an effective method for examining contours over longer timescales and applied to contours in F0 [21, 22].

The smoothing spline with 95% confidence intervals was generated using the *ssanova* function in the gss package [22] in R [23]. Figure 1, plotted using the ggplot2 package [24], shows the results. The results in the two conditions are considered significantly different if the confidence intervals do not overlap.

As shown in Fig. 1, in general, there are significant differences between statement and question in the HC group, and the F0 level of the statement is higher than that of question. In the PD group, however, the differences between two sentence types are not as significant as that in the HC group, and most confidence intervals overlap. Secondly, the F0 level of the PD is lower than the HC, and interrogative F0

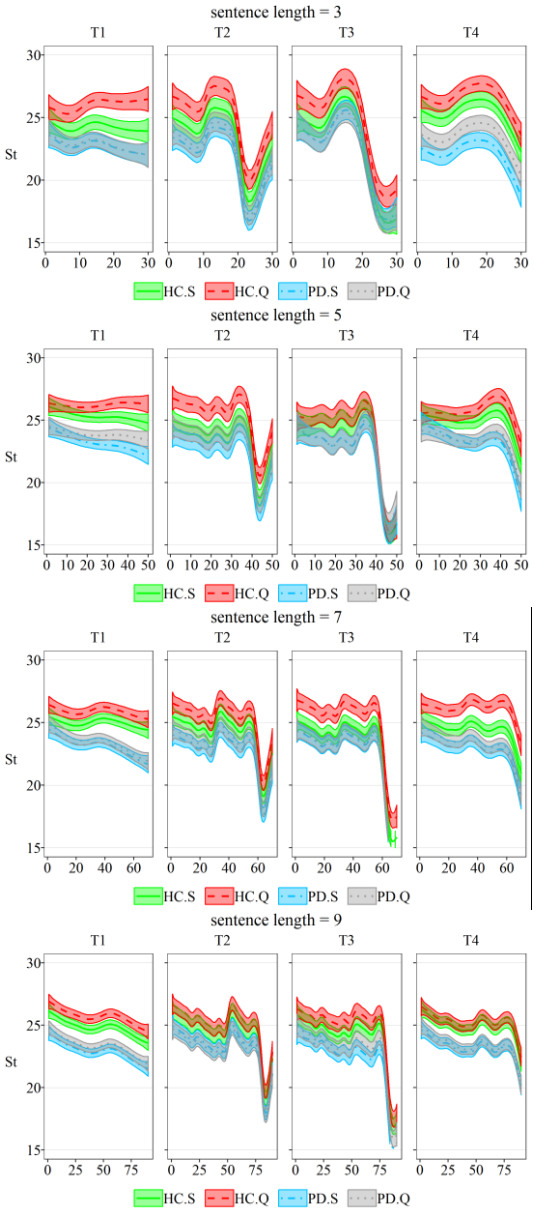

Figure 1: *SS-ANOVA plots of modeled F0 contours with 95% confidence intervals (HC.S represents the statement of HC group, HC.Q represents the question of HC group, etc.).*

level of the PD even lowers than the HC's statement. Finally, the sentence length and final-syllable tone have an influence on the sentence type of two groups. For example, for the HC group, the F0 contrasts between two sentence types are

inconspicuous in 9-syllable sentences ending with T4; while for the PD group, the F0 difference is more obvious than other conditions in 3-syllable sentences ending with T4.

3.2. LMM for acoustic parameters

As mentioned above, the final syllable may present different characteristics from non-final part and our materials vary only in the final-syllable tone. So the parameters were divided into two parts to investigate the global (non-final) and final contrasts between two sentence types. Besides, the intensity ratio of the first syllable to the final syllable was chosen because of the reported difference in intensity contour of questions and statements, a larger intensity ratio was expected for statements than for questions [17]. All the parameters are as follows:

- F0_global: mean F0 of the non-final syllables (St)
- F0_final: mean F0 of the final syllable (St)
- slope_global: F0 slope of the non-syllable utterance
- slope_final: F0 slope of the final syllable
- dur_global: mean duration of the non-final syllables (s)
- dur_final: mean duration of the final syllable (s)
- int_ratio: the ratio of intensity of the first syllable to the final syllable

The Linear Mixed-effects Models (hereafter LMM) for each parameter were conducted using lme4 package [25] in R [23]. As fixed effects, group, sentence type, sentence length, and final-tone (with interaction term) entered into the model. As random effects, we had intercepts for subjects, as well as by-subject random slopes for the effect of sentence type, sentence length, and final-syllable tone. The item was not included as random effects on account of the convergence problem. P-values were obtained by *anova()* function in lmerTest [26], and Satterthwaite's method was used to estimate the degree of freedom. *emmeans()* function in emmeans package [27] was applied to conduct the Tukey's pairwise comparison. High-level interaction terms with the factor of group and sentence type were given priority to reporting, because the significance of low-level interaction and main effects may be distorted by interference in that case.

As for F0_gloabl, there is a significant interaction effect of group * sentence type ($F(1, 38.7) = 14.9$, $p < 0.001$). Figure 2 shows the boxplot of F0_global for the two groups in different sentence types, and the significance levels of simple effects are marked in the figure. Question has a significantly higher F0_gloabl than statement in the HC group ($t(38.7) = 7.5$, $p < 0.001$), while in the PD group there is no significant difference. The HC group is generally higher than the PD group, but there is no significant difference between the two groups in statement or in question, which does not coincide with the result of SS-ANOVA in Fig. 1. This may be because the random effects (especially by-subject effects) in the LMM model explains some variance that belongs to inter-group processing effects in ANOVA.

For F0_final, the model shows a significant interaction effect between group, sentence type, and sentence length ($F(3, 1063) = 2.8$, $p = 0.039$). As shown in Fig. 3, for the HC group F0_final is significantly higher in question than in statement for all sentence lengths, while for the PD group there is a significant difference only in 3-syllable sentences, and the question are generally higher than the statement ($t(237.9) = 2.3$, $p = 0.02$). No significant difference in F0_final is found between the two groups.

For dur_global, the interaction of group*sentence type* sentence length is significant ($F(3, 1025) = 4.2$, $p = 0.006$). Questions are significantly shorter than statements in 3-syllable sentences for the PD group, and in 5- and 9-syllable sentences for the HC group. Figure 4 shows that HC is generally shorter than PD, but there is no significant difference in either statement or question.

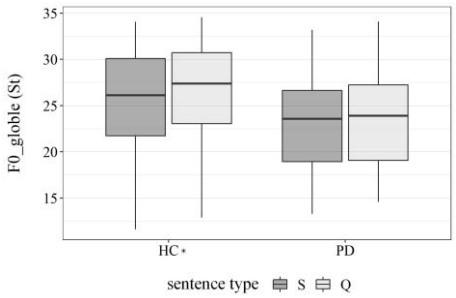

Figure 2: *F0_global boxplot with simple effects significance for two groups in different sentence type (Note: HC* means two sentence types show a significant difference in the HC group, * $p < 0.05$, the same below).*

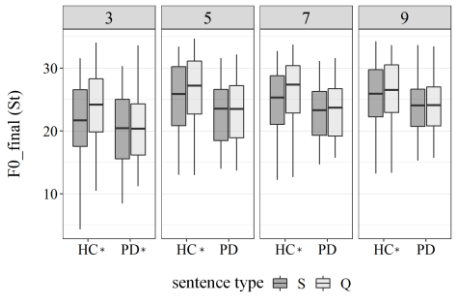

Figure 3: *F0_fianl boxplot with simple effects significance for two groups in different sentence types and sentence lengths.*

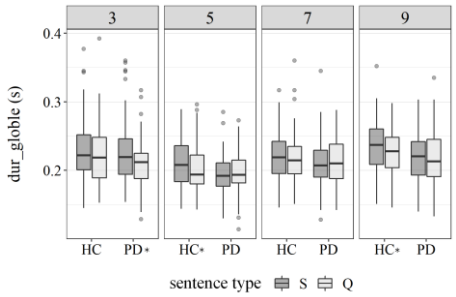

Figure 4: *Dur_global boxplot with simple effects significance for two groups in different sentence types and sentence lengths.*

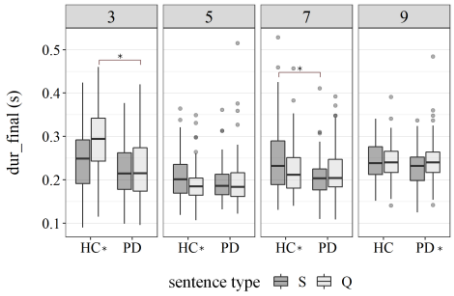

Figure 5: *Dur_final boxplot with simple effects significance for two groups in different sentence types and sentence lengths.*

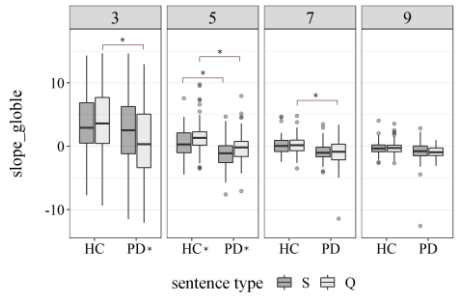

Figure 6: *Slope_global boxplot with simple effects significance for two groups in different sentence types and sentence lengths (Note: 22 outliers outside 3 times SD were removed).*

For dur_final, the interaction effect of group * sentence type * sentence length is significant ($F(3, 1025.2) = 11.9$, $p < 0.001$). Figure 5 shows that for the HC group, the significant difference between two sentence types can be found in 3-, 5-, and 7-syllable sentences. In the 3-syllable sentences, questions are longer than statements, but the result is the opposite in the other cases. For the PD group, in 9-syllable sentences, questions are significant longer than statements ($t(430.4) = 2.3$, $p = 0.024$). Note that the HC group is generally longer than PD, especially in 3-syllable sentences and declarative 7-syllable sentences, which are significant ($p < 0.05$) and the main effect of group is also significant ($F(1, 38) = 5.15$, $p = 0.029$).

For slope_global, there is a significant interaction effect of group * sentence type * sentence length ($F(3, 1039) = 4.35$, $p = 0.005$). The significant difference between the two sentence types is more often on shorter sentences, i.e., the 3-syllable sentence for the PD group and 5-syllable sentence for the HC and PD group. The slope_global of the HC group is larger than the PD group generally, and in 3-syllable and 7-syllable declarative utterances and 5-syllable sentences, the difference between two groups is significant ($p < 0.05$).

For slope_final, group shows a significant main effect ($F(1, 38) = 16.35$, $p < 0.001$), and the PD group (-11.9) is smaller than the HC group (-4.7), indicating that the PD group ends with the steeper declining F0 contour. As for int_ratio, neither the effects of group factor nor interactions with group factor are significant, but the main effect of the sentence type is significant ($F(1, 38.5) = 8.137$, $p = 0.007$). As predicted, statements have the larger slope_final than questions.

4. Discussion

In the first place, the two groups were compared on F0. Although SS-ANOVA showed that the PD group produced lower F0 contours than the HC group, LMM analysis did not find a significant difference in the mean F0 between the two groups. This seemed to coincide with the reports in [11, 12] but conflicted with [8-10]. To make a more reliable conclusion, the future study needs to be conducted on a larger sample.

Nevertheless, it is more meaningful to look at the difference between statement and question among the two groups. For the HC group, questions have significantly higher F0 level than statements, which is consistent with previous studies [4, 5]. In contrast, for the PD group, there is no significant difference in F0 between statement and question. In other words, Mandarin speakers with PD do not raise the overall F0 level to achieve the question, and this poor performance differs from previous studies for German [10] and Cantonese [17] speakers with PD. This suggests that for PDs, the global F0 planning is difficult, especially when tone is concerned. Weakened laryngeal control and reduced amplitude root in PD probably account for this.

The PD group has shorter syllable duration (i.e. faster speech rate) than the HC group, especially in the final syllable, suggesting that PDs' articulatory control degraded at the end of an utterance. Statement are longer than questions for two groups in the non-final syllable, similar to previous studies [5, 7]. For final syllable, sentence length has an impact on the HC group instead of the PD group. For the HC group, the question has longer final-syllabic duration than the statement for 3-syllable sentences, but shorter duration for five and 7-syllable sentences. The PD group shows longer final duration for questions. It may suggest that duration or speech rate is just the incidental characteristic or supplementary cue for intonation contrasts.

The PD group has a steeper slope than the HC group, for both question and statement. Like the faster speech rate, it gives the impression of hurrying to finish the utterances. This is also a reflection of PDs' reduced laryngeal control and motor planning ability.

5. Conclusions

In summary, globally higher F0 is the most important cue for question intonation in Mandarin, and the speakers with PD are flawed in this respect so that unable to differentiate the two sentence types. In addition, PD speech shows faster speech rate and steeper F0 slope, all of which root in the reduced laryngeal control and motor planning ability. In this study, the defects of PD speech are exposed in more complex speech tasks (i.e., sentence intonation). It suggests that more complex speech tasks should be adopted in the scales for PD and more types of speech samples should be included in the automatic evaluation system of PD speech.

6. Acknowledgements

This work is supported jointly by the National Social Science Fund of China (13&ZD189), and the project for Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences (2017STD006). The authors thank Fan Ping, Li Shanpeng and Guan Haitao for their helpful discussions.

7. References

[1] P. Brodal, *The Central Nervous System: Structure and Function*, 4th Edition. New York: Oxford University Press, 2010.

[2] S. Skodda, W. Grönheit, and U. Schlegel, "Intonation and speech rate in Parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission", *Journal of Voice*, vol. 25, no. 4, pp. e199–e205, 2011.

[3] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria", *Journal of Speech and Hearing Research*, vol. 12, no. 2, pp. 246–269, 1969.

[4] X. N. Shen, *The Prosody of Mandarin Chinese*, vol. 118. Berkeley: University of California Press, 1990.

[5] J. Yuan, *Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling*. New York: Cornell University, 2004.

[6] F. Liu and Y. Xu, "Parallel encoding of focus and interrogative meaning in Mandarin intonation", *Phonetica*, vol. 62, no. 2–4, pp. 70–87, 2005.

[7] J. Shen, "A discussion on the pattern of Chinese intonation", *Linguistic Research*, vol. 4, pp. 16–24, 1992. (In Chinese)

[8] G. J. Canter, "Speech characteristics of patients with Parkinson's disease: I. Intensity, pitch, and duration", *Journal of Speech and Hearing Disorders*, vol. 28, no. 3, pp. 221–229, 1963.

[9] J. Illes, E. J. Metter, W. R. Hanson, and S. Iritani, "Language production in Parkinson's disease: Acoustic and linguistic considerations", *Brain and language*, vol. 33, no. 1, pp. 146–160, 1988.

[10] J. K. Ma and R. Hoffmann, "Acoustic Analysis of Intonation in Parkinson's Disease", in *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[11] A. Ghio *et al.*, "F0 characteristics in Parkinsonian speech: contrast between the effect of hypodopaminergy due to Parkinson's disease and that of the therapeutic delivery of L-Dopa", *Revue de laryngologie-to control-rhinologie*, vol. 135, no. 2, pp. 63–70, 2014.

[12] W. Gu, P. Fan, and W. Liu, "Acoustic analysis of Mandarin speech in Parkinson's disease with the effects of levodopa", In: Q. Fang et al. (Eds.), *Studies on Speech Production* (LNAI 10733, Springer), pp. 211–224, 2018.

[13] V. L. Hammen and K. M. Yorkston, "Speech and pause characteristics following speech rate reduction in hypokinetic dysarthria", *Journal of Communication Disorders*, vol. 29, no. 6, pp. 429–445, 1996.

[14] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease", *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 350–367, Jan. 2011.

[15] M. De Letter, P. Santens, M. De Bodt, P. Boon, and J. Van Borsel, "Levodopa-induced alterations in speech rate in advanced Parkinson's disease", *Acta Neurologica Belgica*, vol. 106, no. 1, p. 19, 2006.

[16] R. J. Holmes, J. M. Oates, D. J. Phyland, and A. J. Hughes, "Voice characteristics in the progression of Parkinson's disease", *International Journal of Language & Communication Disorders*, vol. 35, no. 3, pp. 407–418, Sep. 2000.

[17] J. K.-Y. Ma, T. L. Whitehill, and S. Y.-S. So, "Intonation contrast in Cantonese speakers with hypokinetic dysarthria associated with Parkinson's disease", *Journal of Speech, Language, and Hearing Research*, 2010.

[18] J. K. Ma, V. Ciocca, and T. L. Whitehill, "Effect of intonation on Cantonese lexical tones", *The Journal of the Acoustical Society of America*, vol. 120, no. 6, pp. 3978–3987, 2006.

[19] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. 2006. [Online]. Available: http://www.fon.hum.uva.nl/praat/

[20] C. Gu, *Smoothing Spline Anova Models*, 2nd Edition, vol. 297. New York: Springer, 2013.

[21] S. Y. Yiu, "Intonation of statements and question in Cantonese English: Acoustic evidence from a smoothing spline analysis of variance", presented at the 18th International Congress of Phonetic Sciences, Glasgow, Scotland, 2015.

[22] C. Gu, "Smoothing spline ANOVA models: R package gss", *Journal of Statistical Software*, vol. 58, no. 5, pp. 1–25, 2014.

[23] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019.

[24] H. Wickham and W. Chang, "An implementation of the grammar of graphics", 2016. [Online]. Available: http://ggplot2.org.

[25] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4", *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[26] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models", *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[27] R. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*. 2019. [Online]. Available: https://cran.r-project.org/web/packages/emmeans/index.html