Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Individual variability in prosodic marking of locally ambiguous sentences

Clara Huttenlauch[1], *Marie Hansen*[1], *Carola de Beer* [1,2], *Isabell Wartenburger* [1], *Sandra Hanne* [1]

[1]Cognitive Sciences, Department of Linguistics, University of Potsdam, Germany
[2]Faculty of Linguistics and Literary Studies, University of Bielefeld, Germany
c.huttenlauch@web.de, {marie.hansen|isabell.wartenburger|hanne}@uni-potsdam.de, carola.de_beer@uni-bielefeld.de

Abstract

The German case marking system contains case syncretisms, which, along with a relatively free word order, can lead to sentences with local ambiguities, for instance, SVO and OVS sentences with string-identical noun phrases (NPs) in sentence-initial position. Prosodic marking constitutes one possibility for ambiguity resolution. Perception studies showed that listeners are sensitive to f0-manipulations on NP1 of such sentences (e. g., different pitch accents). Here, we investigated which prosodic cues speakers use when they are asked to provide disambiguation in their productions as early as possible. We elicited productions of German SVO and OVS verb-second main clauses, that begin with a case-ambiguous NP1 and are string-identical up to the post-verbal unambiguous NP2. We focused our analysis on the f0-contours in the ambiguous part of the sentences. Overall, there was no consistent f0-pattern that distinguished SVO from OVS sentences. However, analyses with Generalised Additive Mixed Models revealed distinctive f0-contours on the individual level with later and higher f0-peaks on NP1 in SVO vs. OVS sentences. We found that (at least some) speakers systematically distinguish word order in locally case-ambiguous structures by prosodic cues (f0, silent intervals). The variability in our data suggests to consider the individual level when dealing with specific tasks.

Index Terms: speech production, locally ambiguous sentences, word order, fundamental frequency, disambiguation, prosody, individual variability

1. Introduction

In German, despite its rich morphological case system for marking grammatical function, the surface form of noun phrases (NPs) can be ambiguous. For NPs involving feminine and neuter nouns, respectively, the surface form is identical in nominative and accusative case. Case is marked on the determiner: *die* for feminine NPs and *das* for neuter NPs both in nominative as well as in accusative case, respectively. Such NPs are considered case-ambiguous. Furthermore, German allows for a relatively free word order: In addition to subject-verb-object (SVO) sentences, the non-canonical word order of object-verb-subject (OVS) is also possible. Thus, if the determiner *die* or *das* is part of an NP at the beginning of a sentence, the syntactic function of that NP as well as its thematic role remains open: it is ambiguous between subject and direct object as well as between agent and patient. Therefore, the word order configuration could potentially be both, SVO or OVS. If the ambiguity gets resolved at later points in the sentence (e. g., by a case-marked post-verbal NP or by verb inflection), the sentence is called temporarily or locally ambiguous (see (1) and (2)). Besides morphological case markers, prosody, verb semantics, and (visual) context can resolve or influence such thematic role-assignment ambiguities.

For comprehension of spoken locally ambiguous sentences, studies have reported a strong SVO bias in German [1, 2]: Listeners expect NP1 to be the subject, thus agent, and NP2 to be the object, thus patient, of the sentence. Using eye tracking, Weber and colleagues [3] showed, however, that prosody can weaken the SVO bias in comprehension. SVO sentences were presented with a L*+H pitch accent on NP1 and a H* nuclear pitch accent on the verb, while OVS sentences had a L+H* nuclear pitch accent on NP1 (the accents are coded following the GToBI model [4, 5]). As the default position for the nuclear accent in West-Germanic languages is the last argument of the verb [6], in this case NP2, both sentences had a non-default nuclear accent placement [3]. In a similar vein, for case-unambiguous SVO and OVS sentences, Henry and colleagues [2] reported an additive use of morphological and prosodic cues. In their eye-tracking experiment, the presence of both, morphological and prosodic cues, facilitated prediction of the upcoming structure observed by an increase in speed and accuracy in looks to a target object. Furthermore, 5-year-old children have been shown to rely on prosody to determine thematic roles in ambiguous sentences [7]. In summary, studies on language comprehension showed that listeners can make use of prosodic cues (e. g., manipulated f0-contours) for thematic role-assignment (which can serve for disambiguation of locally ambiguous sentences) even before unambiguous morphological cues are accessible [3]. Moreover, prosodic cues increased speed and accuracy of prediction in language comprehension in the presence of morphological cues [2].

For production, [3] reported variability in the f0-contours produced on locally ambiguous OVS sentences with intonation phrase breaks and silent intervals following NP1. It remains open, however, whether there is a clearly differential pattern in f0-contours for SVO and OVS sentences within and across participants or whether the prosodic marking of such sentences is subject to individual variability. In the present study, we analysed productions of locally ambiguous SVO and OVS sentences with a focus on prosodic cues, mainly f0. Our research questions read as follows:

RQ1: How do speakers use prosodic cues (f0, silent intervals) when asked to provide disambiguation as early as possible in productions of locally ambiguous sentences?

RQ2: Do we find differences in the f0-contours in the ambiguous part of the sentence (on NP1 and the following verb) between SVO and OVS sentences within and across speakers?

We focused our analysis on the f0-trajectory in the ambiguous part of the sentences, as comprehension studies have manipulated f0 in this region. We expected large inter-individual variability in the productions, since the task was rather specific and variability especially in productions of OVS structures has been reported previously (e. g., [3]).

 10.21437/SpeechProsody.2022-34

2. Methods

2.1. Participants

Sixteen native speakers of German (12 female, 4 male; mean age 24 years, *SD*: 3.1, age range: 20–30 years) were included in the production study. Two additional speakers took part in the study, but were discarded due to artefacts in the recordings. All participants received course credits or monetary reimbursement for their participation and were naïve to the purpose of the study. The procedure of this study has been approved by the Ethics Committee of the University of Potsdam (approval number 72/2016) and each speaker gave written consent to participate. All speakers had normal hearing defined as an average pure-tone audiometry of 25 dB HL or better for 500, 1000, 2000, and 4000 Hz in the better ear assessed using an audiometer (Hortmann DA 324 series) and calculated following the classification of hearing impairment by the WHO as reported in [8].

2.2. Materials

Items consisted of locally ambiguous and semantically reversible German verb-second main clauses (see (1) and (2)) and corresponding black-and-white line drawings depicting agent, patient, and the action (see Figure 1). We used 20 different transitive verbs and each verb as well as its NPs appeared in two word order conditions: (1) with subject-verb-object word order (SVO) and (2) with object-verb-subject word order (OVS). Thus, a total of 40 items was composed.

(1) Das Kamel tritt nun den Tiger.
the.NOM/ACC camel kicks now the.ACC tiger
'The camel now kicks the tiger.'

(2) Das Kamel tritt nun der Tiger.
the.NOM/ACC camel kicks now the.NOM tiger
'The tiger now kicks the camel.'

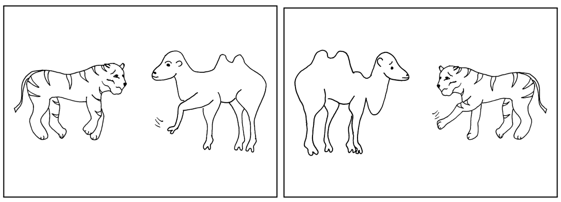

Figure 1: *Example of black-and-white line-drawings used as visual stimuli for SVO (left) and OVS (right) word order.*

All sentences had the same structure: NP1, verb, adverb *nun* (Engl. *now*), NP2. NP1 always contained the neuter determiner *das* which is case-ambiguous between nominative and accusative case. NP2 contained the case-unambiguous masculine determiner *der* in nominative and *den* in accusative case. The determiner of NP2, thus, morpho-syntactically disambiguated the syntactic and thematic roles of both NPs as either the subject/agent or the object/patient of the action. Animate nouns of three different categories were used in subject and object position: persons/humans, animals, and fairy tale characters. Only nouns of the same category were used within the same sentence. The nouns in both positions (NP1 and NP2) were balanced for their mean frequency (DELEXDB, dlexdb.de), their number of phonemes, and number of syllables. The nouns in NP1 can be divided into three categories on the basis of syllable structure and stress pattern: (a) monosyllabic nouns (e. g., HUHN, Engl. *hen*), (b) disyllabic nouns with penultimate stress (e. g., MAEDchen, Engl. *girl*, capital letters refer to stressed syllable), and (c) disyllabic nouns with ultimate stress (e. g., phanTOM, Engl. *phantom*). The semantic reversibility of the thematic roles in the sentences, that is, an equal likelihood for both nouns of the sentence to take the role of agent or patient of the action, was assessed in a rating study on the reversibility of the sentences. Furthermore, comprehension agreement for the pictures was evaluated in a sentence-picture-matching task (for more details see https://osf.io/gychu/). This procedure was similar to the one described in previous studies (e. g., [1]). For conducting the experiment in a semi-interactive setting, two different interlocutors (contexts) were created that were virtually present during the task: a young and an elderly female native speaker of German (the same type of contexts has been used in [9]).

2.3. Experimental procedure

Productions were elicited by means of a referential communication task. Speakers were seated in front of a wide screen (resolution 1920 x 1200) in a sound-attenuated recording booth at the University of Potsdam wearing a headset HSC 271 (AKG Acoustics by Harman, www.akg.com) with over-ear headphones and a condenser microphone. On the screen, speakers were presented first with an audio-visual impression of the corresponding interlocutor in two short videos (one video with a personal introduction of the interlocutor and another one with instructions for the task). The young interlocutor (baseline) was always presented first and the elderly interlocutor in a second block. Each trial started with a fixation cross on the screen and was replaced after 1000 ms by the target and the foil picture with the corresponding sentences printed below the pictures. After a preview time of 4000 ms, the target picture was highlighted with a green frame along with the auditory presentation of the question *Was sehen Sie?* ('What do you see?') via headphones. The question was asked by the interlocutor and was intended to trigger the production (i. e., to simulate question-answer dyads) and as a reminder of the current interlocutor's identity. The task was to produce the target sentence in a way that would allow the interlocutor to identify the target picture "as rapidly and accurately as possible"; the speakers were told that the interlocutor would see the same pictures (without the frame around the target picture) and had to find the matching one. Productions were recorded via an Alesis io12 interface (at 48 kHz). The experiment was run from a Dell laptop using the software Presentation [10]. Speakers produced each item twice, once addressing the young and once the elderly interlocutor.

2.4. Data treatment

In total, 1280 sentences were produced (20 verbs * 2 word order conditions * 2 interlocutor contexts * 16 speakers). In the present analysis, we focused on the productions in the baseline context (n=640). In case the productions contained hesitations (i. e., slips of the tongue, restarts), the part with the hesitation was cut out if the remaining part constituted a fluent utterance (n=6). For three productions (0.5%) the hesitations could not be cut out and the complete productions were excluded from the subsequent analyses of f0.

In the remaining 637 productions, constituent boundaries (NP1, verb + adverb, NP2) were segmented in Praat [11] following standard segmentation criteria [12]. Silent intervals preceding stops were segmented as part of the following constituent and labelled as closure, while silent intervals preceding non-stops were segmented between constituents and labelled as pauses. F0-values were extracted with a customised praat script combining the procedures of Mausmooth [13] and ProsodyPro [14]. For each soundfile, a pitch object was created and unreliable pitch points were removed manually. The pitch contour was smoothed with a bandwidth of 10 Hz before interpolation of pitch points and with a bandwidth of 15 Hz afterwards following the procedure in [13]. Twenty f0-values per constituent (only the intervals that contained speech were considered) were extracted and converted into semitones (st) relative to 1 Hz following [15] in order to ease comparison independent of pitch height. The smoothed f0-contours were plotted with centering at the onset of the verb separately for each speaker and item (SVO and OVS production on top of each other). In a separate step, interval durations were extracted automatically using a Praat script. We collapsed durations of intervals labelled as pauses and closures. The durations of the silent intervals preceding the verb (following NP1) and preceding NP2 (following *nun*) were plotted separately for speaker and word order condition in violin plots. F0-contours and silent intervals were analysed separately.

3. Analyses & Results

In accordance to our rather open research questions, we chose an exploratory analysis procedure consisting of a combination of visual and statistical inspection of the data in several steps. All data and code are available on OSF: https://osf.io/gychu/.

3.1. F0-contours

We started with a visual inspection of the f0-contours within speakers. For most speakers and items, f0-contours of SVO and OVS sentences overlapped quite neatly, thus, showing no visible difference between word order conditions in the ambiguous part of the sentence. Yet, for some speakers, consistent differences in the f0-contours between conditions were noticeable. Across speakers, different f0-contours were produced.

To statistically corroborate the observations from visual analysis in f0-contours between SVO and OVS sentences, we fitted Generalised Additive Mixed Models (GAMMs, [16, 17, 18, 19]) in R [20] using the R package *mgcv* [16, 21]). GAMMs allow to model time-varying data with non-linear patterns controlling for random-effects and autocorrelation [17, 19, 22, 23] and have been successfully used in previous analyses of f0-contours [22, 23, 24].

Since we were interested in distinctive f0-contours for condition within speakers, we decided to fit an individual model for each speaker. The f0 time series (in st, 60 measurements per production, i. e., 20 per constituent) were entered as response variable to the model and word order (as ordered factor) as a predictor (parametric difference term). To directly test whether the difference between SVO and OVS sentences is significant, we added a reference and a difference smooth separately to the model (s(time, bs="tp", k=10) + s(time, by=condition.ord, bs="tp", k=10)) [18, 19, 23]. We compared the model to a nested model without the difference terms using the function compareML() in the R package *itsadug* [25] to check whether their inclusion was justified. Model complexity was increased stepwise with model comparisons at each step. The number of basis functions was doubled from k=10 (default) to 20 if the function gam.check() revealed a low *p*-value and a k-index <1. The final models all included an AR(1) error model to correct for autocorrelation in the residuals (cf. [19, 22]) and a random smooth for item (s(time, item, bs="fs", m=1)).

For significance testing and interpretation of the effect of word order condition on the produced f0-contours, we checked the summary statistics and the difference curves plotted with the plot_diff() function of the R package *itsadug* [25]. Visualisation is crucial for interpreting results of GAMMs [19, 23]. The difference curve visualises the comparison between the non-linear smooths of the two condition levels (here SVO minus OVS) with a pointwise 95% confidence interval (CI). Values above zero indicate larger f0-values in SVO and values below zero indicate larger f0-values in OVS. The difference between the two conditions is significant if the pointwise 95% CI of the difference curve does not include zero. Across all items and speakers, NP2 started on average 0.66 s (*SD* 0.16) after verb onset. We, therefore, consider the time window between the onset of NP1 and 0.66 s (onset of NP2) as a rough approximation of the ambiguous part of the sentence. Note, that for some areas of the utterance, the fitted values are less reliable. This is the case at the beginning and at the end of the utterance (productions differed in their duration, leading to fewer and less reliable f0-measures at the outer extremes; utterance-final glottalisation) and at the onset of the verb (as silent intervals labeled as pauses were dismissed in the extraction of f0-values and the interpolated contour might be disturbed). We are aware that interpretation needs caution as the report of significant intervals of any minimal duration [23] and the decision where to look for differences strongly influence the results [26].

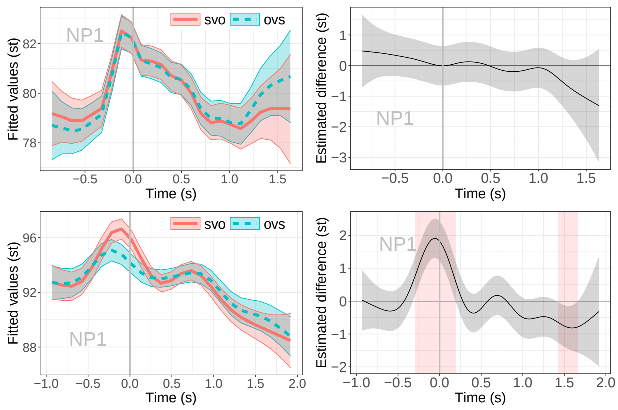

Figure 2: *f0-values (st) predicted by the GAMMs for SVO (red solid line) and OVS (blue dashed line) (left) and predicted differences (SVO-OVS, right) for two speakers (top and bottom row). Time (s) is centered to the offset of NP1/onset of the verb (vertical line). Shaded bands indicate pointwise 95% confidence intervals (CI). The difference is significant if the CI excludes zero (indicated by red shading).*

Across speakers, the difference curves of individual speakers showed a diverse pattern. Two examples of difference curves are given in Figure 2 (right panels) together with the fitted values (left panels) for (a) a speaker for whom the model did not reveal a significant difference (Figure 2 top) and (b) a speaker for whom the model revealed a significant difference in two time windows (Figure 2 bottom). Overall, for ten speakers, the respective models revealed significant differences between SVO and OVS word order within the time window of syntactic ambiguity (i. e., preceding NP2). Note that these differences do not exclude possible further differences during NP2 (as can be seen in the bottom right panel in Figure 2). For four speakers, no significant differences were attested throughout the whole utterance and, for the remaining two speakers, the estimated differences were located more than 0.6 s after the onset of the verb, thus during the case-unambiguous NP2. A closer look at the regions of significant difference between f0-contours preceding NP2 revealed variability across speakers as to location and duration, including locations we previously discussed as less reliable: For some speakers, the significantly different regions were predicted at the very beginning of the utterance (n=2) and relatively local around the onset of the verb (n=2, both speakers produced silent intervals preceding the verb). For other speakers, NP1 was not included in the significant time window, as contours diverged only after the offset of NP1 within the verb (n=1) and divergence even extended into NP2 (n=1). The time windows of the differences had durations between one hundred and more than nine hundred milliseconds. The difference between conditions was mainly (not exclusively) positive, indicating larger f0-values in SVO than in OVS sentences.

3.2. Silent intervals

With regard to silent intervals, we also started with a visual inspection of the data within speakers. As for the f0-contours, the data contained large variability: For ten speakers, we appraised no production of silent intervals neither preceding the verb nor preceding NP2, while three speakers produced silent intervals in one but not the other position, and three speakers produced silent intervals in both positions.

To statistically corroborate the observations from visual analysis, we used a series of intra-individual Wilcoxon Signed Rank Tests [27], since this non-parametric procedure for comparing two groups with non-independent data points can handle non-normally distributed data (which was the case in our data on silent intervals). We used the corresponding base R function to test for statistical significance of the difference in the durations of the silent intervals between SVO and OVS in both positions of interest in the utterance.

The tests revealed statistically significant differences between the conditions in both positions of interest for one speaker (preceding the verb: V=69, p=0.02; preceding NP2: V=2, p=0.03), and in the position preceding NP2 solely for another speaker (V=131, p=0.01). For both speakers, the mean duration of the silent interval preceding the verb was longer in OVS (0.07 s and 0.12 s) than in SVO (0.01 s and 0.04 s), respectively, while preceding NP2, there was a silent interval in SVO (0.02 s) but none in OVS for the first speaker.

4. Discussion

With respect to RQ1, the results reveal that most of our speakers did not use the investigated prosodic cues (f0-trajectories, presence/durations of silent intervals between constituents) distinctively at all in SVO vs. OVS sentences. Only a minority of speakers used f0 and/or silent intervals. The prosodic marking of OVS sentences with longer preverbal silent intervals compared to SVO sentences is partly in line with Weber and colleagues [3] who reported silent intervals for OVS sentences in the productions of individual speakers in their pilot recordings. Overall, the investigated prosodic cues in our study show large inter-individual variability that impedes general conclusions. However, the fact that, intra-individually, the speakers were very stable in their prosodic realisations supports accounts which point towards the importance of individual patterns [28].

Concerning RQ2, intra-individual analyses indicated that there are statistical differences in the f0-contours between SVO and OVS sentences, suggesting that some speakers in our study indeed prosodically disambiguated the two word order conditions early in the sentence (i. e., before the disambiguating NP2) by means of f0. Nevertheless, the majority did not prosodically disambiguate at all or only later in the sentence. Two speakers produced a clearly higher f0-peak on NP1 in SVO vs. OVS sentences. For analysis, GAMMs were fitted on the f0-values of all sentences irrespective of syllable structure and stress pattern. We are, hence, rather cautious in drawing further conclusions about the alignment of the tonal movement with the segmental structure. Clearly, more detailed analyses are necessary in order to allow fine-grained comparisons of our data to previously used stimuli in comprehension studies (such as [2, 3]; for details on two distinct f0-contours see https://osf.io/gychu/).

The finding of large variability between speakers is in line with our expectations. However, we did not expect to find such a high degree of intra-individual consistency in the prosodic contours of locally ambiguous SVO and OVS sentences. One possible reason could be individual difficulties of some speakers to get the non-canonical OVS-reading of the sentences (as reported besides the recordings). Furthermore, since the studied sentences contained only temporary ambiguities which were resolved at the post-verbal NP, there was, strictly speaking, no necessity to resolve the ambiguity via other means (e. g., prosody). This is in no way to say that the speakers did not make an effort to prosodically distinguish SVO from OVS sentences early. The difficulty of the task might be reflected by a paused speaking manner, introducing silent intervals between constituents; a pattern which was apparent in the productions of several speakers. However, the durations of these silent intervals were overall within the range of mean durations of German stops [29] and their difference between SVO and OVS sentences only revealed statistical significance for two speakers.

We do not claim that the investigated prosodic cues are the only means to achieve prosodic disambiguation of locally ambiguous sentences. Future research should examine additional cues (e. g., constituent duration, timing and alignment of f0-movement, intensity) and include the perception side.

5. Conclusions

We analysed productions of locally case-ambiguous sentences with SVO and OVS word order, in which the syntactic ambiguity gets morphologically resolved by the case-unambiguous post-verbal NP2. We explored whether speakers prosodically distinguish the two word order conditions already earlier in the sentence, that is already during the morpho-syntactically ambiguous region. We focused our analysis mainly on f0 as this cue was reported to facilitate disambiguation in comprehension. For most speakers, we found no differences in prosodic cues between SVO and OVS sentences. Nevertheless, since some individual speakers produced systematically distinctive f0-contours, silent intervals, or both, we argue for considering the individual level in order to acknowledge variability in prosodic realisations of the syntactic structure of sentences.

6. Acknowledgements

We thank L. Junack for the drawings and her help with segmenting the data and A. Hofmann for her help with testing participants and annotation. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 317 633 480 - SFB 1287.

7. References

[1] S. Hanne, F. Burchert, R. De Bleser, and S. Vasishth, "Sentence comprehension and morphological cues in aphasia: What eye-tracking reveals about integration and prediction," *Journal of Neurolinguistics*, vol. 34, pp. 83–111, 2015.

[2] N. Henry, H. Hopp, and C. N. Jackson, "Cue additivity and adaptivity in predictive processing," *Language, Cognition and Neuroscience*, vol. 32, no. 10, pp. 1229–1249, 2017.

[3] A. Weber, M. Grice, and M. W. Crocker, "The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements," *Cognition*, vol. 99, no. 2, pp. B63–B72, 2006.

[4] M. Grice and S. Baumann, "Deutsche intonation und GToBI," *Linguistische Berichte*, vol. 191, pp. 267–298, 2002.

[5] M. Grice, S. Baumann, and R. Benzmüller, "German intonation in autosegmental-metrical phonology," in *Prosodic typology: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 55–83.

[6] D. R. Ladd, *Intonational phonology*. Cambridge: Cambridge University Press, 2008.

[7] T. Grünloh, E. Lieven, and M. Tomasello, "German children use prosody to identify participant roles in transitive sentences," *Cognitive Linguistics*, vol. 22, no. 2, pp. 393–419, 2011.

[8] B. O. Olusanya, A. C. Davis, and H. J. Hoffman, "Hearing loss grades and the international classification of functioning, disability and health," *Bulletin of the World Health Organization*, vol. 97, no. 10, pp. 725–728, 2019.

[9] C. Huttenlauch, C. de Beer, S. Hanne, and I. Wartenburger, "Production of prosodic cues in coordinate name sequences addressing varying interlocutors," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 12, no. 1, pp. 1–31, 2021.

[10] (2018) Neurobehavioural systems. Version 20.1. [Online]. Available: https://www.neurobs.com/

[11] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer*, 1992–2017. [Online]. Available: www.praat.org

[12] A. Turk, S. Nakai, and M. Sugahara, "Acoustic segment durations in prosodic research: A practical guide," in *Methods in empirical prosody research*, S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schließer, Eds. Berlin, New York: Walter de Gruyter, 2006, vol. 3, pp. 1–28.

[13] F. Cangemi, "Mausmooth," 2015. [Online]. Available: http://phonetik.phil-fak.uni-koeln.de/fcangemi.html

[14] Y. Xu, "ProsodyPro – a tool for large-scale systematic prosody analysis," in *Tools and Resources for the Analysis of Speech Prosody*. Aix-en-Provence, France: Laboratoire Parole et Langage, 2013, pp. 7–10.

[15] V. Hazan, O. Tuomainen, and M. Pettinato, "Suprasegmental characteristics of spontaneous speech produced in good and challenging communicative conditions by talkers aged 9–14 years," *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 6, pp. S1596–S1607, 2016.

[16] S. N. Wood, *Generalized additive models: An introduction with R*. Boca Raton, FL: CRC press, 2017, 2nd ed.

[17] H. Baayen, S. Vasishth, R. Kliegl, and D. Bates, "The cave of shadows: Addressing the human factor with generalized additive mixed models," *Journal of Memory and Language*, vol. 94, pp. 206–234, 2017.

[18] M. Sóskuthy, "Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction," *arXiv preprint arXiv:1703.05339*, 2017.

[19] M. Wieling, "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English," *Journal of Phonetics*, vol. 70, pp. 86–116, 2018.

[20] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[21] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *Journal of the Royal Statistical Society: Series B (Statistical Methodolody)*, vol. 73, no. 1, pp. 3–36, 2011.

[22] Y.-Y. Chuang, J. Fon, and R. H. Baayen, "Analyzing phonetic data with generalized additive mixed models," *PsyArXiv preprint: psyarxiv.com/bd3r4*, 2020.

[23] M. Sóskuthy, "Evaluating generalised additive mixed modelling strategies for dynamic speech analysis," *Journal of Phonetics*, vol. 84, p. 101017, 2021.

[24] K. Zahner, M. Xu, Y. Chen, N. Dehé, and B. Braun, "The prosodic marking of rhetorical questions in standard Chinese," in *Proceedings of the 10th International Conference on Speech Prosody 2020 (SP20)*, Tokyo, Japan, 2020, pp. 389–393.

[25] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, "itsadug: Interpreting time series and autocorrelated data using GAMMs," 2020, R package version 2.4.

[26] T. B. Roettger, "Researcher degrees of freedom in phonetic research," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 10, no. 1, pp. 1–27, 2019.

[27] S. Siegel, *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, 1956.

[28] F. Cangemi, M. Krüger, and M. Grice, "Listener-specific perception of speaker-specific production in intonation," in *Individual differences in speech production and perception*, S. Fuchs, D. Pape, C. Petrone, and P. Perrier, Eds. Frankfurt: Peter Lang, 2015, pp. 123–145.

[29] C. Kuzla and M. Ernestus, "Prosodic conditioning of phonetic detail in German plosives," *Journal of Phonetics*, vol. 39, no. 2, pp. 143–155, 2011.