Correlation of prosodic boundary cues in German

Fabian Schubö[1], *Sabine Zerbian*[1]

[1]University of Stuttgart

fabian.schuboe@ling.uni-stuttgart.de, sabine.zerbian@ifla.uni-stuttgart.de

Abstract

This study investigates how durational and F0 cues expressing intonation phrase boundaries are correlated in German. Experimental speech data is analyzed as to a correlation between pause duration, pre-boundary lengthening, and F0 range. The results reveal a negative correlation between pause duration and pre-boundary lengthening. This observation is compatible with the assumption that the durational cues are engaged in a trading relationship to fill a timing slot associated with the prosodic boundary, as has been proposed for English and Swedish in prior studies. Furthermore, we observed a positive correlation between pause duration and F0 range, which shows that these cues are implemented in congruence at the same level of the prosodic hierarchy. Altogether, our findings suggest that prosodic boundary cues should be investigated in combination rather than in isolation.

Index Terms: prosody, intonation, prosodic phrasing, duration, F0, pause, pre-boundary lengthening, correlation, German

1. Introduction

In speech production, prosodic boundaries are expressed by means of various phonetic correlates. These include durational cues such as pre-boundary lengthening and silent pauses as well as F0 cues such as the excursion of the F0 contour to the top or bottom of the speaker's register at the end of a prosodic phrase (realizing edge tones). While numerous studies investigated prosodic boundary cues in isolation (e.g., [1] for German), only few studies addressed if and how they are correlated (e.g., [2,3] for English).

Some studies provide opposing findings regarding the relationship between pre-boundary lengthening and the presence or duration of silent pauses: For example, [2] found that the amount of pre-boundary lengthening and the duration of a pause are positively correlated in English. That is, the longer the duration of a silent pause was, the longer was the duration of the phrase-final material. Later studies, however, found a negative correlation between these boundary cues in Swedish [4,5] and English [3]. That is, the longer the duration of the silent pause was, the shorter was the duration of the phrase-final material. This effect has been accounted for by means of a trading relationship between the two durational cues: The increased duration of the phrase-final material and the period of silence are both used to fill an abstract timing interval at the prosodic boundary, which can be implemented with differing proportions of each cue [3]. Alternatively, it has been proposed that pre-boundary lengthening and pause duration are not correlated, but that a silent pause only occurs if the pre-boundary material cannot be lengthened any further, which is referred to as the Stretchability Hypothesis [6]. Under this assumption, a silent pause is inserted to fill a timing interval in cases where pre-boundary lengthening has already been maximized.

To our knowledge, correlations between durational and F0 cues for prosodic boundaries have not been investigated in much detail (but see [7]). Different types of F0 contours at prosodic boundaries are often associated with different categories in a prosodic hierarchy, such as the intonation phrase (IP) and the intermediate phrase (ip) in systems of Tone and Break Indices (e.g., [8] with regard to German). The F0 excursion has been observed to be relatively larger at boundaries that are relatively higher in the prosodic hierarchy (e.g., [9] for German). A similar effect has been found for pre-boundary lengthening and pause duration, which are longer at relatively stronger prosodic boundaries (e.g., [3,10]). Taken together these findings suggest a positive correlation of the amount of F0 excursion, the amount of pre-boundary lengthening, and the duration of pauses across different types of prosodic boundaries. It remains unclear, however, if F0 excursion and durational cues are also positively correlated if they occur at the same type of prosodic boundary, such as at the IP level.

The present study investigates how phonetic cues associated with IP boundaries are correlated in German. The investigation is restricted to pause duration, final segment duration (reflecting pre-boundary lengthening), and the expansion of the F0 range on IP-final words (reflecting F0 excursions that realize H% edge tones). Against the background provided above, the present study tests the hypothesis that these cues are positively correlated. With regard to the durational cues, we also test the alternative hypothesis that pause duration and pre-boundary lengthening are negatively correlated. This alternative would be compatible with the assumption that the durational cues are engaged in a trading relationship to fill a timing window, as proposed in [3]. Related to these aspects, the Stretchability Hypothesis [6] is addressed (which predicts that a silent pause occurs only under the condition that the lengthening of the word is maximized). The presence of a correlation between pause duration and pre-boundary lengthening would provide evidence against this hypothesis, as PBL should neither increase nor decrease when a pause is present. The hypotheses were tested on a highly controlled data set, which is described below.

2. Methods

2.1. Data

The data analyzed in this study is a subset of the data collected in a production study on the domain of pre-boundary lengthening in German (see [11] for results). The production study involved the elicitation and audio-recording of read speech in a laboratory setting. The target words were trisyllabic proper names. The names were controlled for the position of main word stress (penultimate vs. antepenultimate) and the

presence/absence of a final coda consonant. The subset of the data employed for the present study includes only those target words that have penultimate stress and an open syllable in final position (e.g., *Ramona*) and were realized with a following IP boundary.

In order to elicit IP boundaries, the target words were embedded in ambiguous lists of the type [N1 or N2 and N3], which can be interpreted as involving a right-branching structure [N1 or [N2 and N3]] or a left-branching structure [[N1 or N2] and N3]. The target words were always the middle name in the list (N2). Prior studies showed that the structural ambiguity of such lists can be resolved by a prosodic phrasing pattern that reflects the respective structure (e.g., [1,12,13]). A right-branching structure typically involves a prosodic boundary after N1 whereas a left-branching structure typically involves a prosodic boundary after N2. The target word (N2) was thus expected to be followed by a prosodic boundary in one of the two conditions, but not in the other. The production study by [1] showed that German speakers produce IP boundaries for the prosodic disambiguation of such lists. In the present study, the lists were embedded in a carrier sentence, as exemplified in (1).

(1) Ich werde Karolin oder Ramona und Peter einladen
I will Karolin or Ramona and Peter invite
'I will invite Karolin or Ramona and Peter.'

Six target words with penultimate stress and CV.CV.CV structure were employed, each having [a] as the final vowel. Twelve sets of sentences of the sort given in (2) were created, henceforth referred to as items. Each target word occurred in two different items. For elicitation, the lists were set in boldface and the respective branching structure was indicated by underlining. The underlining in (2a) indicates a left-branching structure and the one in (2b) indicates a right-branching structure. The sentences were preceded by a short context story. We recorded 24 native speakers of German from the Stuttgart area aged between 18 and 24 years. Each subject produced both types of structures (within-subjects design), which yielded a total of 288 productions for each branching structure (12 items x 24 subjects). As will be further detailed below, we only included those productions that involved an IP boundary after the target word. These were almost exclusively realizations of the left-branching structure (2a).

(2) a. Ich werde **Karolin oder Ramona** und **Peter** einladen.
b. Ich werde **Karolin** oder **Ramona und Peter** einladen.
'I will invite Karolin or Ramona and Peter.'

The elicitation procedure involved a communication task (basically following [1]). Before the recording session, the subject was familiarized with the type of sentences and instructed to produce each sentence in such a way as to communicate the indicated structure to the experimenter. During the recording session, the subject and the experimenter sat at a table separated by a shoulder-high screen. The stimuli were presented to the subject one by one on a display screen in pseudo-randomized order. The subject read each presented stimulus silently and then produced the sentence containing the target word. For each stimulus, the experimenter saw the sentence containing the target word twice on a printed list, marked once with left- and once with right-branching structure. The experimenter listened to the subject's production, decided which of the two structures was expressed, and checked a box next to the respective sentence on the list. The subject did not see the experimenter's decision and no feedback was given. The communication task was meant to make the subjects produce disambiguating prosodic cues more reliably, as prior studies showed that speakers produce prosodic cues for disambiguation in a consistent way only if they are needed for communication (e.g., [14,15]). The recording sessions took place in a sound-attenuated booth at the University of Stuttgart and lasted approximately 45 minutes.

2.2. Analysis

The target words in the recorded productions were analyzed with regard to the duration of the final vowel, the presence and duration of a following pause, the presence and type of edge tones, and F0 range. All annotations and automated procedures were performed using the acoustics analysis software Praat [16]. The boundaries of the word-final vowel, the boundaries of all pauses, and the boundaries of the utterance were manually annotated (following the guidelines in [17]). A pause was defined as a period of silence of at least 20 ms. The presence and type of edge tones, representing the prosodic phrase categories, were annotated based on the GToBI system [8]. For the analysis of F0 range effects, the points of minimum and maximum F0 on the target words and the point of minimum F0 on the last two words of each production were detected by an automated procedure. The resulting duration values for the vowels, pauses, and utterances, the maximum and minimum F0 values (Hz), and the edge tone annotations were extracted by an automated procedure.

In order to neutralize inter-speaker variation, the duration and F0 data were normalized by means of speaker-specific reference points. The duration values for the word-final vowel were normalized by means of z-score transformation. The z-scores were computed based on the subject-specific mean duration and standard deviation of the final vowel in the target words produced without a following prosodic boundary (see also [18]). The following transformation was applied to each vowel duration value: $z=(x-\mu)/\sigma$, where x is the absolute value, μ is the mean, and σ is the standard deviation. The pause duration values were normalized with reference to the articulation rate of the respective utterance, measured as average syllable duration. Relative pause duration was computed by dividing the absolute pause duration by the average syllable duration of the respective utterance (see also [19]). The F0 minimum and maximum values detected on the target word were converted into semitones with reference to the F0 minimum on the last two words of the respective utterance. For each production, the maximum F0 range on the target word was obtained by computing the difference between the respective maximum and minimum semitone value.

For statistical analyses, we employed the software environment R [20], using the Performance Analytics package for correlation analyses and the lme4 package [21] for linear mixed effects (LME) regression models. See below for details.

3. Results

3.1. Prosodic boundary types

As a first step, the productions were classified as to the type of prosodic boundary that followed the target word. Table 1

presents the frequency of prosodic boundary types assumed in the GToBI system for the productions from the left-branching and the right-branching condition, respectively. As expected, the vast majority of instances from the left-branching condition involved an IP boundary after the target word (96%) whereas the vast majority of instances from the right-branching condition did not involve a prosodic boundary in this position (99%). Only three instances involved an ip boundary after the target words. As the present study is restricted to an analysis of phonetic correlates at IP boundaries, all instances not involving such a boundary after the target word are excluded from the subsequent analyses. That is, the data set used in the following includes 279 productions (277 from the left-branching and 2 from the right-branching condition).

Table 1: *Frequency of prosodic boundary type after the target word for each branching structure.*

	IP	ip	None	Total
Left-bran.	277 (96%)	3 (1%)	8 (3%)	288
Right-bran.	2 (1%)	0	286 (99%)	288

The frequency of the types of edge tones in the productions that involved an IP boundary after the target word are given in Table 2. In most instances, the IP boundary was produced with an H-% edge tone (87%). An L-% edge tone occurred only in 3 instances (1%). These instances were excluded from the subsequent analyses of F0 range because the F0 excursion reaches into the opposite direction, which possibly affects the size of the F0 range. Thus, a total of 276 tokens were included in the analyses of F0 range.

Table 2: *Frequency of edge tone types in the production with an IP boundary after the target word.*

H-%	H-H%	L-H%	L-%	Total
243 (87%)	22 (7%)	11 (3%)	3 (1%)	279

The IP boundaries after the target word involve a silent pause in 251 instances (90%). Only 28 productions involved an IP boundary without a silent pause (10%). These productions were excluded from the subsequent analyses of pause duration.

3.2. Correlation analysis

A correlation analysis was performed for each combination of IP boundary cues (relative pause duration, z-score-transformed segment duration, and normalized F0 range in semitones), employing Spearman's rank correlation coefficient. The results of the tests are summarized as a correlation matrix in Figure 1, showing the distribution of each variable on the diagonal, the bivariate scatter plots with a fitted line on the bottom of the diagonal, and the correlation values and significance codes on the top of the diagonal. The tests yielded a positive correlation between pause duration and F0 range ($r(249)=0.39$, $p<0.001$) and a negative correlation between pause duration and segment duration ($r(249)=-0.22$, $p<0.001$). There was no significant effect for a correlation between F0 range and segment duration ($r(274)=-0.017$, $p=0.78$).

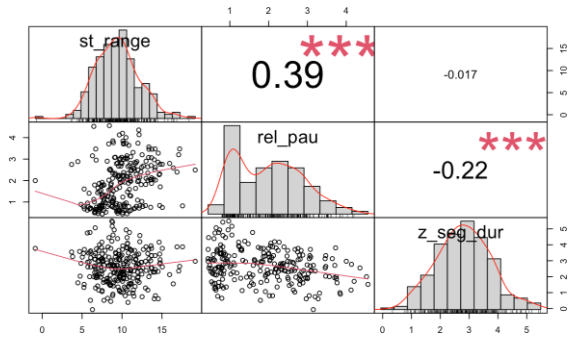

Figure 1: *Correlation matrix for F0 range (st_range), relative pause duration (rel_pau) and z-score-transformed segment duration (z_seg_dur).*

3.3. Short and long pauses

As shown in Figure 2, the histogram for relative pause duration suggests a bi-modal distribution of the data. A Hartigan's dip test yielded a significant effect ($D=0.035$, $p<0.05$), which suggests a deviation from uni-modality. We therefore divided the productions involving pauses into two groups: productions with short pauses and productions with long pauses at the IP boundary. The threshold between long and short pauses was set at the relative duration value 1.5, as pauses near this value have the lowest frequency between the first and the second peak in the distribution. Short pauses occurred in 103 and long pauses in 148 instances.

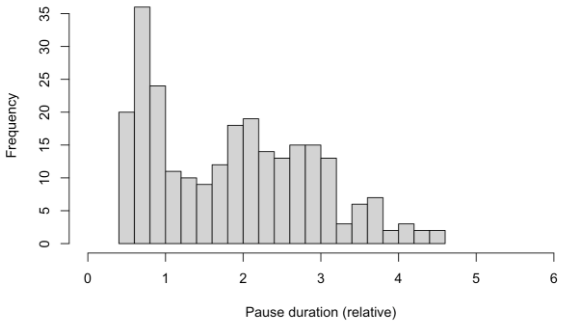

Figure 2: *Histogram for relative pause duration (n=251).*

3.4. Segment duration as a function of pause type

Figure 3 presents the duration data of the IP-final segment (z-score-transformed) by pause type. The boxplots suggest that the segment duration is larger before short pauses than before long pauses. An LME model accounting for segment duration as a function of pause type (levels: short, long) was fitted to the data. The random structure included intercepts for speaker and item. The output is given in Table 3. The model estimated that the duration of IP-final segments is 0.27 points shorter before long pauses than before short pauses. The model was tested against a reduced model without pause type as a fixed factor (thus only involving random factors) by means of a likelihood ratio test, which yielded a significant effect ($\chi^2(1)=4.07$, $p<0.05$). The AIC value of the full model is more than two points lower than

the one of reduced model, which suggests that the full model is significantly better.

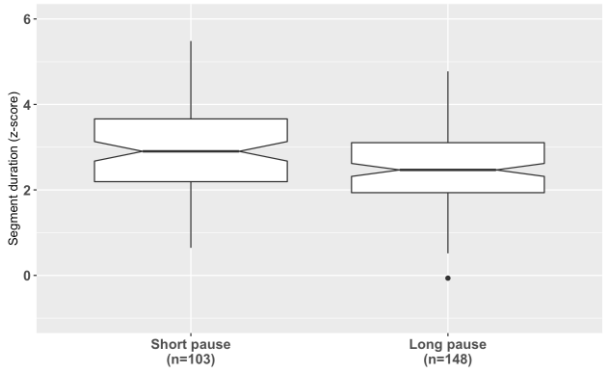

Figure 3: *Duration (z-score) of IP-final segments before short and long pauses.*

Table 3: *Output of an LME model accounting for z-score-transformed IP-final segment duration as a function of pause type.*

	Estimate	Std. Error	t value
Intercept	2.81	0.16	17.82
Pause type *long*	-0.27	0.13	-2.05

3.5. F0 range as a function of pause type

Figure 4 presents the data of the maximum F0 range on the target words (in semitones) by pause type. The boxplots suggest that the F0 range is larger before long pauses than before short pauses. An LME model accounting for F0 range as a function of pause type (levels: short, long) was fitted to the data. The random structure included intercepts for speaker and item. The output is given in Table 4. The model estimated that the F0 range is 1.32 semitones larger on IP-final words before long pauses than before short pauses. The model was tested against a reduced model without pause type as a fixed factor (thus only involving random factors) by means of a likelihood ratio test, which yielded a significant effect ($\chi^2(1)=14.3$, $p<0.001$). The AIC value of the full model is more than two points lower than the one of the reduced model, which suggests that the full model is significantly better.

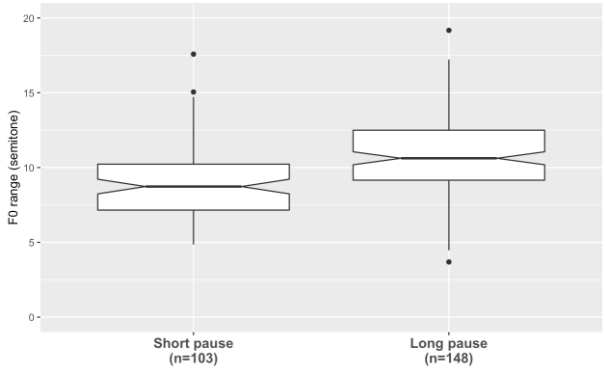

Figure 3: *F0 range (semitones) on target words before short and long pauses.*

Table 4: *Output of an LME model accounting for F0 range (in semitones) as a function of pause type.*

	Estimate	Std. Error	t value
Intercept	9.23	0.42	22.18
Pause type *long*	1.32	0.34	3.87

4. Discussion

The results revealed a negative correlation between pause duration and segment duration (reflecting pre-boundary lengthening) and a positive correlation between pause duration and F0 range (reflecting the degree of excursion in the realization of H% edge tones). There was no correlation between F0 range and pre-boundary lengthening. Pauses were subdivided into short and long pauses based on the frequency distribution of the pause duration data. Pre-boundary lengthening was found to be longer before short pauses than before long pauses while the F0 range was found to be smaller before short pauses than before long pauses.

The findings suggest that pause duration and pre-boundary lengthening are involved in a trading relationship, as found by [3] for English and [4,5] for Swedish. Our findings for German thus contrast with the findings by [2] for English, who observed a positive correlation of the durational cues. Our findings are also incompatible with the Stretchability Hypothesis [6], as a correlation between pre-boundary lengthening and pause duration would be unexpected if the presence of a pause only occurred under the condition that the pre-boundary material cannot be lengthened any further.

The observation that pause duration and F0 range are positively correlated when used for marking IP boundaries shows that a correlation between these cues does not only result from differences between levels in the prosodic hierarchy (e.g., IP vs. ip), but also occurs at prosodic boundaries of the same type. Altogether, the observations from the present study suggest that prosodic boundary cues should be investigated in combination rather than in isolation.

5. Acknowledgements

This study was conducted as part of the research project *Pre-boundary lengthening in a cross-linguistic perspective* at the University of Stuttgart, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 416902968.

6. References

[1] C. Petrone, H. Truckenbrodt, C. Wellmann, J. Holzgrefe-Lang, I. Wartenburger, and B. Höhle, "Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists," *Journal of Phonetics*, vol. 61, pp. 357–366, 2017.

[2] W. E. Cooper and J. Paccia-Cooper, "Syntax and Speech," Harvard University Press, 1980.

[3] F. Ferreira, "Creation of prosody during sentence production," Psychological Review, vol. 100, no. 2, 233–253, 1993.

[4] G. Fant, A. Kruckenberg, and L. Nord, "Prosodic and segmental speaker variations," *Proceedings of the ESCA Workshop on Speaker Characterization in Speech Technology*, pp. 106–120, 1990.

[5] M. Horne, E. Strangert, and Mattias Heldner, "Prosodic boundary strength in Swedish: Final lengthening and silent interval

duration", *Proceedings of the 13th International Congress of Phonetic Sciences*, pp. 170–173, 1995.
[6] E. O. Selkirk, "Phonology and Syntax: The Relation between Sound and Structure," Cambridge: MIT Press, 1984.
[7] G. Kentner, I. Franz, C. A. Knoop, and W. Menninghaus, "The final lengthening of pre-boundary syllables turns into final shortening as boundary strength levels increase," *Journal of Phonetics* 97, 101225, 2023.
[8] M. Grice, S. Baumann, and R. Benzmüller, "German intonation in auto-segmental metrical phonology," In S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation*, Oxford: Oxford University Press, pp. 55–83, 2005.
[9] H. Truckenbrodt, "Upstep and embedded register levels," *Phonology* 19, pp. 77–120.
[10] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1707–1717, 1992.
[11] F. Schubö and S. Zerbian, "The patterns of pre-boundary lengthening in German," in F. Schubö, S. Zerbian, S. Hanne and I. Wartenburger (eds.), Prosodic boundary phenomena. Berlin: Language Science Press, to appear.
[12] A. E. Turk and S. Shattuck-Hufnagel, "Multiple targets of phrase-final lengthening in American English words," *Journal of Phonetics*, vol. 35, no. 4, pp. 445–472, 2007.
[13] G. Kentner and C. Féry, "A new approach to prosodic grouping," *The Linguistic Review*, vol. 30, no. 2, pp. 277–311, 2013.
[14] J. Snedeker and J. Trueswell, "Using prosody to avoid ambiguity: effects of speaker awareness and referential context," *Journal of Memory and Language*, vol. 48, pp. 103–130, 2003.
[15] F. Schubö, A. Roth, V. Haase and C. Féry, "Experimental investigations on the prosodic realization of restrictive and appositive relative clauses in German," *Lingua*, vol. 154, pp. 65–86, 2015.
[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer", Computer program, version 6.1.05, URL http://www.praat.org/, 2019.
[17] A. E. Turk, S. Nakai, and M. Sugahara, "Acoustic Segments Durations in Prosodic Research: A Practical Guide," in A. Steube (ed.), *Language, Context & Cognition – Methods in Empirical Prosody Research*, pp. 445–472, 2006.
[18] B. Peters, K. J. Kohler and T. Wesener, "Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache," in K. J. Kohler, F. Kleber and B. Peters (eds.), *Prosodic Structures in German Spontaneous Speech* (AIPUK 35a), pp. 143–184, 2005.
[19] P. Hansson, "Perceived boundary strength," *Proceedings of the 7th International Conference on Spoken Language Processing*, 2002.
[20] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org, 2019.
[21] D. Bates, M. Mächler, B. Bolker and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 67(1), pp. 1–48, 2015.