Prosodic boundary phenomena

Edited by

Fabian Schubö

Sabine Zerbian

Sandra Hanne

Isabell Wartenburger

Studies in Laboratory Phonology 12

language
science
press

Studies in Laboratory Phonology

Chief Editor: Martine Grice
Editors: Doris Mücke, Taehong Cho

In this series:

3. Roettger, Timo B. Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments.
4. Mücke, Doris. Dynamische Modellierung von Artikulation und prosodischer Struktur: Eine Einführung in die Artikulatorische Phonologie.
5. Bergmann, Pia. Morphologisch komplexe Wörter im Deutschen: Prosodische Struktur und phonetische Realisierung.
6. Feldhausen, Ingo, Jan Fliessbach & Maria del Mar Vanrell (eds.). Methods in prosody: A Romance language perspective.
7. Tilsen, Sam. Syntax with oscillators and energy levels.
8. Ben Hedia, Sonia. Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics.
9. Easterday, Shelece. Highly complex syllable structure: A typological and diachronic study.
10. Roessig, Simon. Categoriality and continuity in prosodic prominence.
11. Schmitz, Dominic. Production, perception, and comprehension of subphonemic detail: Word-Final /s/ in English.
12. Schubö, Fabian, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.). Prosodic boundary phenomena.

ISSN: 2363-5576

Prosodic boundary phenomena

Edited by

Fabian Schubö
Sabine Zerbian
Sandra Hanne
Isabell Wartenburger

language
science
press

Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.). 2023. *Prosodic boundary phenomena* (Studies in Laboratory Phonology 12). Berlin: Language Science Press.

ISBN: 978-3-96110-409-3 (Digital)
978-3-98554-067-9 (Hardcover)

ISSN: 2363-5576
DOI: 10.5281/zenodo.7777469
Source code available from www.github.com/langsci/379
Errata: paperhive.org/documents/remote?type=langsci&id=379

Cover and concept of design: Ulrike Harbort
Proofreading: Agnes Kim, Amir Ghorbanpour, Amy Amoakuh, Brett Reynolds, Daejin Kim, Elliott Pearl, Jeroen van de Weijer, María Milla, Nele Ots, Neneng Sri, Jean Nitzke, Rebecca Madlener, Vadim Kimmelman
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: XƎLATEX

Language Science Press
xHain
Grünberger Str. 16
10243 Berlin, Germany
http://langsci-press.org

Storage and cataloguing done by FU Berlin

Freie Universität Berlin

Contents

Acknowledgments iii

Preface v

1 The patterns of pre-boundary lengthening in German
Fabian Schubö & Sabine Zerbian 1

2 Segmental cues to IP-initial boundaries: Data from English, Spanish, and Portuguese
Ricardo Napoleão de Souza 35

3 Chunking an unfamiliar language: Results from a perception study of German listeners
Nele Ots & Piia Taremaa 87

4 Developmental changes in prosodic boundary cue perception in German-learning infants
Caroline Wellmann, Julia Holzgrefe-Lang, Hubert Truckenbrodt, Isabell Wartenburger & Barbara Höhle 119

5 Age effects on linguistic prosody in coordinates produced to varying interlocutors: Comparison of younger and older speakers
Clara Huttenlauch, Marie Hansen, Carola de Beer, Sandra Hanne & Isabell Wartenburger 157

Index 193

Acknowledgments

This collaboration was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – projects "Preboundary lengthening in a cross-linguistic perspective" (project no. 416902968, GZ ZE 940/3-1) and SFB 1287 project B01 (project no. 317633480).

Preface

In spoken language comprehension, the hearer is faced with a more or less continuous stream of auditory information. Prosodic cues, such as pitch movement, pre-boundary lengthening, and pauses, incrementally help to organize the incoming stream of information into prosodic phrases, which often coincide with syntactic units. Prosody is hence central to spoken language comprehension and is "the skeletal structure on which the rest of the utterance depends" (Frazier et al. 2006: 248). Accordingly, some models assume that the speaker produces prosody in a consistent and hierarchical fashion (e.g., Nespor & Vogel 1986). While there is manifold empirical evidence that prosodic boundary cues are reliably and robustly produced and effectively guide spoken sentence comprehension across different populations and languages, the underlying mechanisms and the nature of the prosody-syntax interface still have not been identified sufficiently. This is also reflected in the fact that most models on sentence processing completely lack prosodic information.

This edited book volume is grounded in a workshop that was held in 2021 at the annual conference of the *Deutsche Gesellschaft für Sprachwissenschaft* (DGfS). The five chapters cover selected topics on the production and comprehension of prosodic cues in various populations and languages, all focusing in particular on processing of prosody at structurally relevant prosodic boundaries.

With respect to the prosodic cues investigated, the different contributions refer to the most common cues crosslinguistically, such as increased segment duration in phrase-initial position (referred to as domain-initial strengthening in Napoleão de Souza 2023 [this volume]) and in phrase-final position (referred to as final lengthening in Ots & Taremaa 2023 [this volume], Huttenlauch et al. 2023 [this volume], and Wellmann et al. 2023 [this volume], referred to as pre-boundary lengthening in Schubö & Zerbian 2023 [this volume]). Pauses and F0-related measures, such as F0-range (also referred to as pitch range), and rises, are also addressed in several studies (Ots & Taremaa 2023 [this volume], Huttenlauch et al. 2023 [this volume], Wellmann et al. 2023 [this volume]). We have refrained from unifying the terminology used across individual chapters, especially in the case of the boundaries themselves. Here, reference is made to prosodic units and domains, prosodic phrases, Intonational Phrases, clause boundaries and breaks, and each contributor defines these terms in the respective chapter as relevant.

Regarding theory and modelling, the contributions by Schubö & Zerbian 2023 [this volume] and Napoleão de Souza 2023 [this volume] deal with the interrelation of prosodic boundaries and lexical prominence. Boundary processes are explored across four languages, and it emerges that the syllable carrying the main stress serves as an anchor for boundary phenomena. Both studies share the reference to the work of Katsika (2016) who showed for Greek that stress on phrase-final words has an effect on boundary-related lengthening but not on phrase-initial lengthening of the following word. In her proposed (gestural) account, lexical and phrasal prosody interact in a systematic and coordinated way at prosodic boundaries.

From a methodological perspective, Schubö & Zerbian (2023 [this volume]), Wellmann et al. (2023 [this volume]), and Huttenlauch et al. (2023 [this volume]) worked with lists of three names which differ in their syntactic branching. Due to these parallels in the stimuli, the results of these studies can inform each other with reference to the Proximity/Similarity Model (Kentner & Féry 2013), specifically as the Principle of Anti-Proximity is concerned in the rendition of the middle name with respect to the strengthening of prosodic cues at its boundary.

With respect to the specific aims, research questions and theoretical contributions of the different studies in this volume, Schubö & Zerbian (2023 [this volume]) investigate the initiation and scope of pre-boundary lengthening on the phrase-final word in German. Native adult speakers read out sentences in which the target word varied with respect to the position of word stress (penultimate vs. antepenultimate syllable) and the presence/absence of an additional segment at the end of the word. In result, pre-boundary lengthening was reliably found on the stressed syllable – and its start was shifted from the onset consonant to the following vowel of the stressed syllable when a coda consonant was added to the words with penultimate stress. This indicates that the scope of pre-boundary lengthening in German is determined by the prosodic structure as well as the segmental composition of the phrase-final words as it has been claimed for other languages as well.

Moving away from prosodic boundary phenomena in a single language, a crosslinguistic comparison of boundary phenomena at the beginning of prosodic units was undertaken by Napoleão de Souza (2023 [this volume]). This chapter compares domain-initial strengthening in three lexical stress languages: English, Spanish, and Portuguese. The study addresses the question of how domain-initial strengthening is expressed acoustically across the three different languages, and how it affects the acoustic properties of segments in fully unstressed syllables in prenuclear post-boundary positions. From the crosslinguistic comparison, the study concludes that acoustic correlates of boundary marking extend beyond the initial segment in unstressed CV syllables (in Spanish and Portuguese the vowel is affected as is the stressed syllable in all three languages).

How prosodic cues can be used by non-native listeners to chunk a speech stream of a language they do not know compared to native speakers of that language has been studied by Ots & Taremaa (2023 [this volume]). To this end, German and Estonian listeners were asked to listen to spontaneous utterances spoken in Estonian and to mark the point in time when they perceived a break between words. While Estonian listeners were guided by clause boundaries marked by longer pauses and intonational rises, German listeners were also sensitive to phrase-final lengthening and intensity drop. This indicates that non-native listeners rely on bottom-up processing for prosodic boundary identification, while native adult speakers also apply their top-down knowledge to chunk the incoming speech stream.

A similar bottom-up processing strategy needs to be applied by newborns and infants in language acquisition. Wellmann et al. (2023 [this volume]) investigate developmental changes in the processing of intonation phrase boundaries which (in German) are mainly characterized by pitch change, final lengthening, and a silent pause. These cues have been shown to have different weightings in perception in different languages. Between the age of six and eight months, infants' prosodic processing undergoes an important development: moving away from the necessity of all of these three cues or a combination of pause and final lengthening for detection of a boundary to a detection based on a combination of pitch and final lengthening without a pause. This shift towards more "independence" from the pause cue reflects a language-specific shift of attention to boundary markings that are functionally relevant in the to-be-learned ambient language.

Finally, Huttenlauch et al. (2023 [this volume]) investigate how the other end of the age spectrum affects the realization of intonation phrase boundaries. Here, the productions of younger speakers of German (Huttenlauch et al. 2021) are compared to older speakers of German while they produced coordinated three-name sequences without and with internal grouping of the first two names. Both age groups marked the grouping globally using all three prosodic cues (i.e., F0 range, final lengthening, and pause) in line with the Proximity/Similarity model by Kentner & Féry (2013). Prosodic grouping was unaffected by age despite age-related longer absolute durations and larger variability in the productions of older participants. Furthermore, as both groups do not or only minimally adapt to different virtual communication partners, the data support models of situational independence of disambiguating prosody (e.g., Speer et al. 2011).

With this volume, we hope to have compiled an interesting compendium on different prosodic boundary phenomena which comprises crosslinguistic evidence as well as evidence from non-native listeners, infants, adults, and elderly speakers, highlighting the important role of prosody in both language production and comprehension.

References

Frazier, Lyn, Katy Carlson & Charles Clifton Jr. 2006. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 10(6). 244–249. DOI: 10.1016/j.tics.2006.04.002.

Huttenlauch, Clara, Carola de Beer, Sandra Hanne & Isabell Wartenburger. 2021. Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Laboratory Phonology* 12(1). 1–31. DOI: 10.5334/labphon.221.

Huttenlauch, Clara, Marie Hansen, Carola de Beer, Sandra Hanne & Isabell Wartenburger. 2023. Age effects on linguistic prosody in coordinates produced to varying interlocutors: Comparison of younger and older speakers. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 157–192. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777534.

Katsika, Argyro. 2016. The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics* 55. 149–181. DOI: 10.1016/j.wocn.2015.12.003.

Kentner, Gerrit & Caroline Féry. 2013. A new approach to prosodic grouping. *The Linguistic Review* 30. 277–311. DOI: 10.1515/tlr-2013-0009.

Napoleão de Souza, Ricardo. 2023. Segmental cues to IP-initial boundaries: Data from English, Spanish, and Portuguese. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 35–86. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777528.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris Publications. DOI: 10.1515/9783110977790.

Ots, Nele & Piia Taremaa. 2023. Chunking an unfamiliar language: Results from a perception study of German listeners. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 87–117. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777530.

Schubö, Fabian & Sabine Zerbian. 2023. The patterns of pre-boundary lengthening in German. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 1–34. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777526.

Speer, Shari R., Paul Warren & Amy J. Schafer. 2011. Situationally independent prosodic phrasing. *Laboratory Phonology* 2(1). 35–98. DOI: 10.1515/labphon.2011.002.

Wellmann, Caroline, Julia Holzgrefe-Lang, Hubert Truckenbrodt, Isabell Wartenburger & Barbara Höhle. 2023. Developmental changes in prosodic boundary cue perception in German-learning infants. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 119–156. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777532.

Chapter 1

The patterns of pre-boundary lengthening in German

Fabian Schubö & Sabine Zerbian
University of Stuttgart

It has been observed for English that pre-boundary lengthening (PBL) is initiated on the syllable with main stress of the phrase-final word. Furthermore, the results of some studies suggest that the scope of PBL varies depending on the segmental composition of the phrase-final word. The present study investigates the scope of PBL in German. We report on a production experiment that tested for the position of word stress (penultimate vs. antepenultimate) and the presence/absence of an additional segment at the end of the word (CV.CV.CV vs. CV.CV.CVC) as predictors for the initiation of PBL. The results revealed that the initiation of PBL occurred on the stressed syllable across conditions. Furthermore, PBL was initiated later when a coda consonant was added to the words with penultimate stress, shifting the initiation point from the onset consonant to the following vowel of the stressed syllable. These observations suggest that the nuclear vowel of the main stress syllable serves as an anchor for PBL, but the initiation occurs earlier if the amount of material between the nuclear vowel and the prosodic boundary is limited. Thus, in line with findings from other languages, the scope of PBL in German is determined by the prosodic structure as well as the segmental composition of the phrase-final word.

1 Introduction

Pre-boundary lengthening (PBL) has been identified as one of the major correlates of prosodic phrasing. That is, segments in phrase-final position are produced with longer duration than the same segments in phrase-medial position. This effect has been attested a stable correlate of boundary production and a reliable cue for boundary perception (e.g., Petrone et al. 2017). The observation that

Fabian Schubö & Sabine Zerbian. 2023. The patterns of pre-boundary lengthening in German. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 1–34. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777526

PBL occurs in various languages, involving different prosodic systems, suggests that it might be a universal phenomenon (see, e.g., Vaissière 1983). At the same time, it has been found that PBL is implemented language-specifically with reference to the given phonological system, which suggests that it must be learnt by speakers (e.g., Nakai et al. 2009 with regard to Northern Finnish).

It has been found that PBL operates primarily on the rime of the phrase-final syllable. Additionally, in some languages, PBL can affect the material of the penultimate syllable and even reach until the antepenultimate syllable of the phrase-final word. The scope of PBL may thus span several syllables. This scope is henceforth referred to as "PBL domain". The extent of the PBL domain is connected to aspects of prosodic prominence. In languages with word stress, it has been observed that PBL is initiated on the last syllable with main stress preceding the boundary (e.g., White 2002 for British English). That is, the domain reaches from the main stress syllable to the end of the phrase-final word. Depending on the location of this syllable, the point of PBL initiation occurs earlier or later in a word so that the PBL domain varies in size. Furthermore, it has been found that the PBL domain is of fixed duration and overlaps with a part of the phrase-final word (e.g., Byrd & Saltzman 2003 with reference to American English). Thus, the initiation of PBL differs as to the number (and type) of given segments immediately preceding the phrase boundary. If material is added to the end of the word, the initiation point is expected to shift further to the right. As for the temporal dynamics of the affected segments, a strong tendency towards a pattern of progressive lengthening has been observed across languages (e.g., Kohler 1983, Silverman 1990 for German; Byrd et al. 2006 for American English). That is, the closer the location of the affected segment is to the prosodic boundary, the larger is the relative amount of lengthening.

The present study investigates the patterns of PBL in German. The aim is to delineate the PBL domain and the temporal dynamics of the affected material. We report on a production experiment that tested if (a) prosodic prominence and (b) the segmental composition of the phrase-final word affect the initiation of PBL.[1] The design controlled for the position of main word stress (penultimate vs. antepenultimate), the composition of the final syllable rime in words with penultimate stress (CV.ˈCV.CV vs. CV.ˈCV.CVC), and the presence/absence of a following prosodic boundary. The results revealed that (a) PBL is initiated on the stressed syllable across conditions, and that (b) the presence of an additional coda

---

[1] A subset of the data gained in this study and a preliminary account were presented in Schubö & Zerbian (2020). This subset includes the productions from 12 out of 24 subjects comprising the target words with penultimate word stress. The chapter at hand presents the full dataset and theoretical account.

consonant in the words with penultimate stress shifts the initiation to the next segment. Furthermore, our findings suggest a weak form of progressive lengthening: The amount of PBL gradually increased by tendency, but also showed an interruption of this pattern in some instances on the material preceding the final rime. The final rime consistently showed an abrupt increase of lengthening on the nuclear vowel (independent of the stress pattern). This supports the view that the final rime has a central role in the implementation of a prosodic boundary by means of PBL (e.g., Turk & Shattuck-Hufnagel 2007).

The paper is structured as follows: The following subsections summarize the relevant background from prior studies on PBL. This is followed by an introduction of the major aspects of German prosody and the statement of the research question and hypotheses. Section 2 details the methods employed in the production experiment and presents the steps in the data analysis. Section 3 presents the results. Section 4 discusses the findings, also addressing crosslinguistic aspects and ends with some concluding remarks.

1.1 The initiation and scope of pre-boundary lengthening

The results from several studies suggest that the initiation of PBL is connected to a specific phonological constituent, such as a syllable or rime. In particular studies on English have provided evidence for the assumption that the initiation of PBL occurs on the last main stress syllable preceding the prosodic boundary (White 2002, Turk & Shattuck-Hufnagel 2007). This effect has been termed the Word Rime hypothesis, reflecting the observation that the first location of lengthening often occurs on the vowel or coda consonant of the main stress syllable. For example, White (2002) tested words like *SPECtre* and *SPECtacle* in British English and found that PBL occurs on the coda consonant of the main stress syllable independent of its position. Similarly, Turk & Shattuck-Hufnagel (2007) tested words like *MIchigan* and *JaMAIca* in American English and found that PBL occurs on the vowel of the main stress syllable. Their study also controlled for the presence/absence of an accent on the stressed syllable and found that PBL applies to the main stress syllable independent of the presence of an accent. This suggests that word stress, and not phrasal stress, is the relevant predictor. Several studies on different languages observed a pattern that is compatible with the Word Rime hypothesis (e.g., Kohler 1983 for German; Cambier-Langeveld 1997 for Dutch; Krull 1997 for Estonian; Cambier-Langeveld 2000 for British English; Nakai et al. 2009 for Northern Finnish). However, the results from these studies do not provide independent evidence for this pattern, as alternative explanations could also account for the data (see below for Kohler's findings on German). Furthermore, it has been found that PBL can be initiated earlier in words with pre-final stress than in words with final stress, but the initiation point is not necessarily located on the stressed syllable (Katsika 2016 for Greek).

As for German, most studies that investigated PBL only tested for an effect on the final syllable (e.g., Peters et al. 2005, Schubö et al. 2015, Petrone et al. 2017). To our knowledge, only two prior studies addressed the extent of the domain of PBL in German: Kohler (1983) analysed acoustic speech data from two speakers who produced the indefinite pronouns *eine* [ˈaɪ.nə] (‘one’) and *einige* [ˈaɪ.ni.gə] (‘some’) in utterance-medial and utterance-final position, respectively. He found that the initiation of PBL occurred on the stressed syllable in both words. This is compatible with the Word Rime hypothesis; yet, as the main stress syllable is initial in both of these words, it could also be the case that PBL operates on the entire prosodic word in German (see also Silverman 1990 and Turk & Shattuck-Hufnagel 2007 for this point). Moreover, it is unclear whether PBL also affects material preceding the utterance-final word. In order to link the initiation of PBL to the main stress syllable, we must exclude the possibility that the syllable preceding the disyllabic test word *eine* also underwent PBL. Silverman (1990) addressed these problems by comparing the durational patterns in a pair of trisyllabic words that differed only as to the location of main word stress. One of the words comprised penultimate stress (*umLAgern* ‘to besiege’) whereas the other one comprised antepenultimate stress (*UMlagern* ‘to relocate’). In Silverman’s study, two German native speakers were recorded, who produced these words six times in phrase-medial and phrase-final position. Silverman’s analysis of the acoustic speech data revealed that PBL operates on the entire prosodic word in German. The presence of PBL on material preceding the main stress syllable must however be assumed to result from a different factor. Similar to this finding for German, experimental data on American English also revealed that PBL can occur on the antepenultimate syllable of a word with penultimate stress (Cho et al. 2013). The variability found in these languages calls for further investigation of the patterns of PBL.

Some models do not posit a connection between PBL initiation and phonological constituency, but assume that the scope of PBL has fixed duration and overlaps with the phrase-final material (e.g., Byrd & Saltzman 2003, Byrd et al. 2005, 2006). From this point of view, the PBL domain is aligned with the phrase boundary at its right edge whereas its left edge is determined by the phrase-final material in terms of the number and intrinsic length of given segments. In the framework of Articulatory Phonology (e.g., Browman & Goldstein 1992, Goldstein et al. 2006), this pattern has been accounted for as resulting from a clock-slowing gesture, the so-called π-gesture (e.g., Byrd & Saltzman 2003, Byrd et al. 2005, 2006), which slows down the articulatory movements at the end of a prosodic phrase and thus leads to a lengthening effect. Such models entail that the initiation of PBL depends on the number of segments overlapped by the PBL domain or π-gesture (depending on the theoretical account). This predicts that PBL is initiated at a later point in a word that has an additional consonant in the final syllable coda than in the same word without the additional consonant. For example, we would expect a later point of PBL initiation in the word *bananas* (comprising a plural suffix) than in the word *banana* (without the plural suffix). We will refer to this assumption as the Overlap hypothesis.

In some languages, both phonological constituency and the type of phrase-final segments have an impact on the scope of PBL. For example, it has been found for Dutch that PBL mainly occurs on the rime of the final syllable; however, in case the rime comprises a vowel that is not expandable, such as a schwa, the initiation point occurs on preceding material (Cambier-Langeveld 1997). A combination of phonological structure and segmental composition has also been observed in Japanese: Seo et al. (2019) found that in disyllabic words PBL is initiated on the vowel of the penultimate syllable as long as it does not contain a coda consonant. In disyllabic words consisting of two CVN syllables, the initiation point occurs on the coda consonant of the penultimate syllable, which can be understood as a shift induced by additional phonetic content. Yet, the authors also attested an impact of the word prosodic structure: In case the words bore a lexical pitch accent anchored to the initial syllable, there was no effect of PBL on the final syllable.

These findings on the PBL domain are inconsistent in several ways. This particularly applies to findings on American English: For example, while Turk & Shattuck-Hufnagel (2007) observed that PBL has a large scope reaching until pre-final main stress syllables, Byrd et al. (2006) found that the scope is limited and that there is no interaction with stress. Furthermore, Byrd & Riggs (2008) observed that the initiation of PBL was shifted to pre-final stressed syllables only by one out of three subjects. Furthermore, Cho et al. (2013) observed the presence of PBL on the antepenultimate syllable in a word with penultimate stress. These inconsistencies call for further research on the initiation of PBL with reference to the position of main word stress. The reason for these inconsistent findings might result from differences in the methods used in prior studies, as the studies differed with regard to aspects such as stimuli, type of data collection, and number of participants.

1.2 The amount and distribution of pre-boundary lengthening

Various studies observed a pattern of progressive lengthening towards the phrase boundary (e.g., Kohler 1983, Silverman 1990 for German; Berkovits 1994 for Hebrew; Byrd et al. 2006 for American English; Nakai et al. 2009 for Northern Finnish; Seo et al. 2019 for Japanese). That is, the amount of PBL progressively increases from one segment to the next in the PBL domain, so that the effect is strongest on the final segment. This pattern might be affected by the expandability potential of specific segments; for example, it has been found that oral stops involve a lower amount of PBL than other types of consonants in American English (Klatt 1976), Hebrew (Berkovits 1993a,b), and Dutch (Hofhuis et al. 1995). It has also been found that, once initiated, PBL can be interrupted on intermediate elements (Cambier-Langeveld 1997 for Dutch; Turk & Shattuck-Hufnagel 2007 for American English). For example, Turk & Shattuck-Hufnagel (2007) found that American English words with antepenultimate stress involved PBL on the rime of the stressed syllable and on the rime of the final syllable, but not on the intermediate material. They also observed "a weaker version of progressive lengthening" (2007: 459), which entails that the amount of PBL globally increases from left to right, but this increase can be interrupted locally, resulting in a lower amount of PBL on a segment in comparison to the amount of the prior segment (thus, it progresses with a "medial dip").

The progressive lengthening pattern often involves a comparatively large increase of lengthening on the phrase-final syllable, leading to a larger slope of progressive lengthening in this position (e.g., Klatt 1975, Kohler 1983, Berkovits 1994, Turk & Shattuck-Hufnagel 2007, Seo et al. 2019). For example, Kohler (1983) found that German words with pre-final stress involve a considerably larger amount of PBL on the final syllable (87–176%) than on the penultimate syllable (15–31%). The data from some production studies suggest that the large amount of increase occurs on the rime of the final syllable: For example, in their study on American English, Turk & Shattuck-Hufnagel (2007) observed 15 percent of lengthening on the onset of the final syllable followed by 71 percent of lengthening on the following syllable rime (mean percentages based on the data from four subjects). Furthermore, Seo et al. (2019) argue that the final rime constitutes the major unit for the distribution of PBL in Japanese, showing, among other things, that the amount of lengthening on the rime of an open syllable (CV) is comparable to the amount on the rime of a closed syllable (CVN).

1.3 German prosody

Before turning to the present study, we will briefly outline the prosodic properties of German. German prosody closely resembles the prosodic system of English. Syllables may be open or closed and can contain single consonants or consonant clusters both in onset and in coda position. With a few exceptions, consonant clusters comply with the Sonority Sequencing Principle (SSP), that is, the degree of sonority decreases towards the edges of a syllable (see, e.g., Selkirk 1984 for this principle). Voiceless alveolar and post-alveolar obstruents can occur at the peripheries of a consonant cluster in violation of the SSP. Voiced coda obstruents undergo devoicing. The rhotic is usually vocalized as [ɐ] in coda position. Several phonotactic constraints apply, including the prohibition of the glottal fricative and the palatal glide in coda position (see, e.g., Hall 1992 for an overview).

Word stress is assigned to either of the last three syllables in a morphologically simple word; yet, polysyllabic words with final stress are rare in German. In words with three or more syllables, there is a tendency for penultimate stress if the penultimate syllable is closed, and for antepenultimate stress if the penultimate syllable is open (Wiese 1996). According to Delattre (1965), German exhibits a tendency towards word-initial stress, but trisyllabic words do not statistically differ as to the frequency of penultimate and antepenultimate stress. The most prevalent phonetic correlate of word stress in German is duration: Vowels and consonants exhibit longer duration in stressed than in unstressed syllables (e.g., Dogil & Williams 1999).

Phrasal stress is assigned with respect to syntactic structure, rhythmic patterns, information structural conditions, and other meaning-related aspects (see, e.g., Truckenbrodt 2006). Phrasal stress is realized by a pitch accent aligned with a main stress syllable as well as by longer duration. Different systems employing Tone and Break Indices (ToBI) are offered in the literature (e.g., Grice et al. 2005, Peters 2018). For the annotation of tonal events, the present study adopts the system proposed in Grice et al. (2005), referred to as German Tone and Break Indices (GToBI). This system assumes a set of six pitch accents (L*, H*, L+H*, L*+H, H+L*, H+!H*), two phrase tones (L-, H-), and two boundary tones (L%, H%). Nuclear stress is usually assigned to the rightmost phrasal stress position and implemented by means of a pitch accent with relatively larger prominence than the preceding ones in the utterance. The nuclear pattern at the end of a prosodic phrase (i.e., the last pitch accent in combination with the following phrase and/or boundary tone) may express specific pragmatic meanings (see, e.g., Grice et al. 2005). For example, a pattern involving (L+)H* H-(%) is often employed for expressing incompleteness whereas a pattern involving L+H* L-% is often employed for expressing a contrastive assertion (Grice et al. 2005: 71).

Two levels of prosodic phrasing are distinguished in GToBI, referred to as Intonational Phrase (IP) and intermediate phrase (ip), respectively. The former involves a relatively stronger and the latter a relatively weaker prosodic boundary at its right edge. The prosodic boundaries on both levels can be expressed by means of boundary tones, PBL and pauses (see Petrone et al. 2017 for a study on the production and perception of these cues). The pitch movements induced by boundary tones can involve rising, falling, or falling-rising patterns on the material between the last pitch accent and the end of the phrase. In utterance-medial position, they usually involve a rising or falling-rising pattern, whereas in utterance-final position they usually involve a falling pattern (see, e.g., Truckenbrodt 2002, 2007). As stated above, PBL initiation was found on the last main stress syllable preceding the boundary (Kohler 1983), but there is also some evidence for lengthening of the prior syllable (Silverman 1990). The amount of lengthening has been found to increase progressively towards the end of the prosodic phrase (Kohler 1983, Silverman 1990).

1.4 Research questions and hypotheses

The present study investigates the patterns of PBL in German speech production, addressing the question of what determines the initiation of PBL. Specifically, it is tested if the initiation of PBL is affected by (a) the position of main word stress and/or (b) the number of segments in the phrase-final word. The respective hypotheses are stated in (1). The statement in (1a) captures the Word Rime hypothesis, which predicts that PBL begins on the last main stress syllable before the prosodic boundary. Thus, if this hypothesis holds, it is expected that words with different stress positions differ with regard to the point of PBL initiation and the scope of the PBL domain. Attesting this pattern for German would strengthen the assumption that prosodic prominence serves as a predictor for PBL initiation in languages with a stress-based prosodic system. The statement in (1b) is in compliance with the Overlap hypothesis, which entails that the scope of PBL is of fixed duration and overlaps with a portion of the phrase-final word. Thus, if this hypothesis holds, it is expected that additional material at the end of the word leads to a shift of PBL initiation to a later point, such as the following segment or syllable. It is also possible that both hypotheses hold, in which case the initiation of PBL would change in accordance with the position of main word stress and at the same time would shift to a later position if additional material is present at the end of the word.

(1) a. PBL is initiated on the nuclear vowel of the main stress syllable and persists until the end of the phrase-final word (Word Rime hypothesis).
    b. The initiation of PBL is delayed if a coda consonant is added to the final syllable (Overlap hypothesis).

Furthermore, this study addresses the relative amount of lengthening among the segments affected by PBL. According to the Progressive Lengthening hypothesis (2), it is expected that the amount is relatively larger on segments that are relatively closer to the end of the prosodic phrase; however, given prior findings from other languages (see §1.2), this pattern might not be applied consistently so that the relative amount of lengthening locally decreases or lengthening is completely absent in intermediate positions.

(2) The amount of PBL progressively increases towards the end of the prosodic phrase (Progressive Lengthening hypothesis).

The predictions were tested by conducting a production experiment, which is reported on in the next section. The experiment involved the elicitation and audio-recording of read speech in a laboratory setting. Given the inconsistencies found in prior studies (see §1.1), we chose to employ a carefully controlled design.

2 Methods

2.1 Stimuli

The stimuli employed in the production experiment were controlled for the position of main word stress, the presence/absence of a final coda consonant, and the presence/absence of a prosodic boundary. We employed two types of target words, which were both trisyllabic proper names. The first type comprised CV.ˈCV.CV structure, involving penultimate word stress (e.g., RaMOna). These words were elicited under two conditions affecting the final rime: In one condition, they were in accusative case and retained their structure. In the other condition, they were in genitive case and comprised a suffix *-s*, which is implemented as a voiceless alveolar fricative in the word-final coda, yielding a CV.ˈCV.CVC structure (e.g., RaMOnas). The second type of target words comprised antepenultimate main word stress (e.g., KArolin). These words varied with regard to the presence of a coda consonant in the penultimate and/or final syllable. Given that proper names with antepenultimate stress and the same internal syllable structure as the words with penultimate stress are rare in German, we decided to also include words that deviated with regard to the presence of coda consonants.

Prosodic boundaries after the target words were elicited by means of lists of the type [N1 or N2 and N3], which can be interpreted as comprising a left-branching structure [[N1 or N2] and [N3]] or a right-branching structure [[N1] or [N2 and N3]]. The target words were in position N2. Prior studies showed that speakers disambiguate such lists by means of prosodic phrasing, inserting a boundary after N2 in the left-branching case and after N1 in the right-branching case (e.g., Kentner & Féry 2013, Petrone et al. 2017, Huttenlauch et al. 2021 for German; Wagner 2005, Turk & Shattuck-Hufnagel 2007 for English; see also Huttenlauch et al. 2023 [this volume] and Wellmann et al. 2023 [this volume]). The lists were medially embedded in carrier sentences. The sentences were preceded by a short context story. The branching structure was indicated by setting the list in italics and underlining its sub-constituents. An example item is given in (3). The target word of the pair in (3a) involves penultimate stress and lacks a final coda consonant (*Ramona*). The first sentence comprises a right-branching structure, which renders the target word in phrase-medial position, and the second sentence comprises a left-branching structure, which renders the target word in phrase-final position. The target word in (3b) involves penultimate stress and a final coda consonant (*Ramonas*). In this case, the sentence comprises an elliptic right-node-raising construction. Finally, the target word in (3c) involves antepenultimate stress (*Karolin*). Here, the structure of the sentences is the same as in (3a), but the first two names are exchanged, so that the name with antepenultimate stress occurs in N2 position.

(3) a. Ich werde *Karolin oder Ramona und Peter* einladen.
Ich werde *Karolin oder Ramona und Peter* einladen.
'I will invite Karolin or Ramona and Peter.'
b. Ich werde *Karolins oder Ramonas und Peters Freunde* einladen.
Ich werde *Karolins oder Ramonas und Peters Freunde* einladen.
'I will invite Karolin's or Ramona's and Peter's friends.'
c. Ich werde *Ramona oder Karolin und Peter einladen.*
Ich werde *Ramona oder Karolin und Peter* einladen.
'I will invite Ramona or Karolin and Peter.'

The context stories consisted of three to four sentences, as illustrated in (4). The story in (4a) preceded the sentences in (3a) and (3c). Since the object in (3b) had a different structure, the context story was slightly modified for reasons of coherence, as illustrated in (4b).

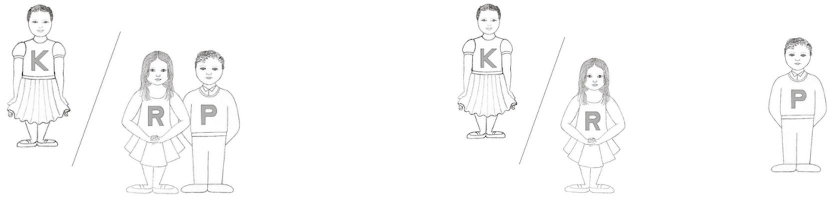
(4) a. Max feiert bald seinen Geburtstag. Er hat bereits seine besten Freunde eingeladen. Nun überlegt er, wen er noch einladen soll, Max denkt:

‘Max will soon celebrate his birthday. He already invited his best friends. Now he is wondering who else he could invite. Max is thinking:’

b. Max feiert bald seinen Geburtstag. Er hat bereits seine besten Freunde eingeladen. Nun überlegt er, auch noch deren Freunde einzuladen. Max denkt:

‘Max will soon celebrate his birthday. He already invited his best friends. Now he is considering to invite their friends as well. Max is thinking:’

In order to facilitate the interpretation of the lists, pictures with drawings of persons grouped according to the constituent structure were presented below the target sentences. Figure 1 illustrates the pictures for the sentences in (3a). The persons were marked with the initial letter of the respective name in the list (here, *Karolin*, *Ramona*, and *Peter*). The picture in (a) represents the right-branching structure, where the target word is in phrase-medial position, and the one in (b) represents the left-branching structure, where the target word is in phrase-final position.



(a) right-branching structure (b) left-branching structure

Figure 1: Examples of pictures showing persons grouped according to the constituent structure of the lists. Ramona (R) is the target.

2.2 Design, subjects and procedure

As described above, the design involved six conditions (2 phrase positions × 2 stress positions + 2 coda conditions). We employed six names with penultimate and six names with antepenultimate stress as target words and created twelve items of the sort presented in §2.1 (each name occurred in two different items). This yielded a total of 72 target expressions, which are given in Appendix A. Furthermore, we employed 90 filler expressions. The stimuli were pseudo-randomized and presented to the subjects in a within-subject design. We audio-recorded 24 native speakers of German from the Stuttgart area aged between 18 and 25 years. This yielded 288 productions per condition (12 items × 24 subjects) and 1,728 productions in total (288 × 6 conditions). The recording sessions took place in a sound-attenuated booth at the University of Stuttgart and lasted 44 minutes on average across subjects. The recordings were made digitally with a Sennheiser ME 62 microphone and stored on hard disk in Waveform Audio File Format (mono sound, 44.1 kHz sampling frequency, 16-bit resolution).

The recording session was self-controlled by the subject using a computer mouse. During the session, the subject and the experimenter were sitting at a table, separated by a shoulder-high screen. The stimuli were presented to the subject one by one on a display screen. The subject was instructed to first read the context and the target sentence silently and then to decide which of the two interpretations was indicated by the underlining of the constituents and the picture given below. After that, the subject started a recording phase of six seconds by clicking on a button on the display screen and then read the target sentence out loud. If the subject was not satisfied with the production, he/she could repeat it after clicking on the recording button again. In this case, the recording of the prior production was deleted. The subject was allowed to repeat the production of a sentence as often as he/she wanted to. After the recording, the subject mouse-clicked a different button on the display screen in order to move on to the next stimulus. The software used for this procedure was written in Python by the first author.

Furthermore, the elicitation procedure involved a communication task (similar to Petrone et al. 2017). The subject was instructed to produce the sentences in such a way that the experimenter could understand which of the two branching structures was expressed. The experimenter saw a printed list with both alternatives for each sentence and had to assign the production to one out of the two alternatives by checking a box on the list. The subject did not see the experimenter's decision and no feedback was given. The experimenter identified the correct structure in 97 percent of cases across subjects and conditions. This procedure was supposed to make the subjects produce the disambiguating prosodic cues more reliably, as it has been found that speakers use prosodic cues for disambiguation in a consistent way only when they are aware of a need for disambiguation (e.g., Snedeker & Trueswell 2003, Schubö et al. 2015). We acknowledge that this procedure might have elicited a focus intonation pattern, as the participants were required to communicate one out of two possible structures. Given that lengthening is also a correlate of focus in German (e.g., Féry & Kügler 2008), we cannot exclude the possibility that the participants produced focus-related lengthening in addition to PBL.

Preceding the recording session, the subjects were familiarized with the type of sentences and the ambiguity involved. They saw an example for each branching structure (including the respective underlining of the constituents and the picture) and read a short text describing the meaning difference. At the beginning of the recording session, the subjects produced five sentences that were not part of the test material, but involved the same type of ambiguity and indication of structure. These productions did not enter the analysis and were deleted after the recording session.

In a prior study on German (Petrone et al. 2017), it was found that an experiment design such as the one employed in the present study leads to a consistent production of IP boundaries (rather than ip boundaries). We chose to employ a design of this type in order to consistently elicit IP boundaries. Designs with less control might cause variation in the type of prosodic boundaries. Given that the amount of PBL is expected to be larger at relatively stronger boundaries (e.g., Peters et al. 2005), such a variation would be problematic.

2.3 Pre-analysis: Pitch accents and boundary tones

A pre-analysis of the intonation patterns on the target words was performed using the GToBI system (Grice et al. 2005). This was applied in order to verify that the productions were consistent with regard to phrasal prominence on the target words and that they involved the expected phrasing patterns. The first author (who is highly familiar with the intonation patterns of German and the GToBI system) manually annotated each production as to the presence and type of pitch accent on the target word. Furthermore, the presence/absence and type of prosodic boundary immediately following and immediately preceding the target word was annotated by the first author. The types of pitch accent and boundary tones were identified based on the local shape of the F0 contours, the global F0 pattern, and the auditory impression of the tonal event. In 30 productions, the target word was unaccented (19 from the right-branching and 11 from the left-branching condition). These tokens were excluded from the subsequent analyses, as the absence of an accent might have affected segment duration. Furthermore, 40 productions (39 from the right-branching and 1 from the left-branching condition) involved a prosodic boundary immediately preceding the target word (i.e., after the conjunction *oder* ‘or’). These tokens were also excluded from the subsequent analyses, as phrase-initial strengthening might have affected the duration of the word-initial segments. Altogether, 1,658 productions were included in the subsequent analyses, which corresponds to 96 percent of the collected tokens.

Table 1 shows the frequency and type of prosodic boundaries realized under the right- and left-branching condition, respectively, indicated by the boundary tone types from the GToBI system. As expected, the majority of productions from the right-branching condition did not involve a prosodic boundary after the target word whereas the majority of productions from the left-branching condition did involve a prosodic boundary in this position. In the latter case, the vast majority of instances ended with an H% boundary tone, indicating the presence of an Intonation Phrase (IP) boundary with a continuation rise at the right edge. Also, there were 20 productions from the right-branching conditions that did involve a prosodic boundary after the target word (16 with H%, 3 with H-, and 1 with L%; see Table 1) and 15 productions from the left-branching condition that did not involve a prosodic boundary after the target word. These productions were included in the subsequent analyses and treated as phrase-medial or -final in accordance with the presence/absence of a boundary tone after the target word. Thus, the categorization of the productions was based solely on their surface prosodic pattern and not on the condition under which they were elicited.

Table 1: Frequency of boundary tone types in the right-branching and left-branching condition

Boundary tone	H%	H-	L%	L-	none
Right-branching	16	3	1	0	**786**
Left-branching	**821**	2	10	4	15

The GToBI annotations revealed that the vast majority of boundary tones signal an IP boundary (i.e., H% and L%). This is in line with the findings by Petrone et al. (2017), who used a similar design and observed a consistent use of IP boundaries. Our data is thus largely consistent with regard to the phrasing level, which avoids variation in PBL resulting from different types of boundaries.

Table 2 presents the frequency of pitch accent types on the target word in phrase-medial position (productions not involving a boundary tone after the target word) and phrase-final position (productions involving a boundary tone after the target word). The most frequent type of pitch accent in phrase-medial position was H* (n = 500; 62 percent), followed by L+H* (n = 191; 24 percent), and L*+H (n = 92; 11 percent). In phrase-final position, the most frequent type was L+H* (n = 567; 66 percent), followed by L*+H (n = 241; 28 percent). Only few monotonal pitch accents occurred in this position. Thus, the target words in the different phrase positions involve different tendencies as to the distribution of pitch accent types. We acknowledge that these tendencies might constitute a confound that could affect the duration of specific segments in the target words (in particular in the stressed syllable and the following one). Testing for a correlation of pitch accent type and segment or syllable duration is beyond the scope of this study and should be addressed in future research. Moreover, inter-speaker differences with regard to these aspects should be explored.

Table 2: Frequency of pitch accent types in phrase-medial and phrase-final position

Pitch accent	H*	L*	L+H*	L*+H	unclear
Phrase-medial	500	12	191	92	6
Phrase-final	23	25	567	241	1

The penultimate stress words with and without a final coda consonant (e.g., *Ramona* vs. *Ramonas*) were elicited by means of different syntactic structures: The words with a final coda consonant were part of an elliptic right-node raising construction, which was not the case for the words without a final coda consonant. In order to check if the different syntactic constructions might have induced different pitch accent patterns, we compared the frequency of the most common pitch accent types (H*, L+H*, and L*+H) realized on the words with penultimate stress with and without a final coda consonant in phrase-medial and -final position, respectively. Table 3 indicates that both conditions show the same tendencies with regard to the distribution of pitch accent types for each phrase position. In phrase-medial position, the most common type in both conditions is H*, followed by L+H*. Only relatively few instances of L*+H occur in this position. The words with a final coda consonant were less often produced with an H* and slightly more often produced with an L+H*, but the relative distribution among the pitch accent types is similar for both conditions. In phrase-final position, the most common type in both conditions is L+H* and fewer instances of L*+H were produced on the words with a final coda consonant, but the relative distribution is similar.

Table 3: Frequency of H*, L+H*, and L*+H in phrase-medial and phrase-final position for the penultimate stress words in elliptic constructions/coda present and non-elliptic constructions/coda absent

	Phrase-medial			Phrase-final		
Pitch accent	H*	L+H*	L*+H	H*	L+H*	L*+H
Coda absent	184	76	12	4	190	72
Coda present	133	89	27	10	220	59

2.4 Analysis

2.4.1 Acoustic speech segmentation

We manually annotated the segment boundaries of the target words based on spectrographic and waveform information, following the guidelines for acoustic speech segmentation provided by Turk et al. (2006). These guidelines suggest that the locations of the segment boundaries should be identified based on abrupt spectral changes caused by the onsets and releases of consonantal constrictions (rather than by voicing criteria). Thus, the segmentation procedure primarily relied on acoustic landmarks that were caused by the consonantal constriction gestures. For example, sibilants were segmented based on the onset and offset of frication energy whereas nasal stops were segmented based on abrupt spectral changes at the points of closure and release, marking an abrupt decrease of energy in comparison to the surrounding vowels. As a secondary cue, the onset and offset of F2 energy was taken into account, which the guidelines suggest particularly with regard to weak fricatives and pre-pausal or utterance-final vowels. For visual inspection and annotation, we employed the acoustics analysis software Praat (Boersma & Weenink 2019). After the annotation process was completed, the duration values of the intervals defined by the identified consonantal constrictions were extracted by means of an automated procedure.

2.4.2 Statistics

For statistical analyses, we employed the software environment R (R Core Team 2018) and the `lme4` package (Bates et al. 2015). Separate linear mixed effects (LME) models were fitted to the data for each type of target word (penultimate stress with coda, penultimate stress without coda, antepenultimate stress). The models account for DURATION as a function of PHRASE POSITION (levels: medial, final) and SEGMENT POSITION (with interaction term). The levels of SEGMENT POSITION included all relevant combinations of the syllable position in the word (antepenultimate, penultimate, final) and the internal syllable structure (onset, nucleus, coda). The interaction term is motivated based on the assumption that the amount of PBL is relatively larger on segments that are closer to the prosodic boundary. As random factors, we included intercepts and slopes for SUBJECT and intercepts for ITEM. Due to non-convergence, the slopes for SUBJECT were removed from the model fitted to the data for the words with antepenultimate stress. Significance between the levels of PHRASE POSITION was tested at each SEGMENT POSITION by using the `multcomp` package (Hothorn et al. 2008). The results are presented below. Model outputs for all coefficients are given in Appendix B.

3 Results

3.1 The scope of lengthening

Figure 2 presents the duration data for the target words with penultimate stress and CV.ˈCV.CV structure (e.g., RaMOna). The light boxes show the data from the productions in phrase-medial position and the dark boxes show the data from the productions in phrase-final position. The codes above the plots indicate the significance level of the p-values obtained by the post-hoc comparisons. The boxplots for the segments in the antepenultimate syllable (C1 and V1) do not suggest a significant difference between the two phrase positions, and the respective comparison did not yield a significant effect (C1: $\beta = 3.6, \mathrm{SE} = 1.9, t = 1.9, p = 0.0598$; V1: $\beta = -0.1, \mathrm{SE} = 1.9, t = -0.5, p = 0.616$). The comparison for C1 was however near significant. The model estimated a longer duration of 4 ms in phrase-final position. The boxplots for the consonants and vowels of the penultimate and final syllable clearly suggest a longer duration in phrase-final position, and the comparisons yielded highly significant effects, respectively (C2: $\beta = 10.7, \mathrm{SE} = 1.9$, $t = 5.6, p < 0.001$; V2: $\beta = 13.9, \mathrm{SE} = 1.9, t = 7.3, p < 0.001$; C3: $\beta = 8.1, \mathrm{SE} = 1.9$, $t = 4.3, p < 0.001$; V3: $\beta = 80.1, \mathrm{SE} = 1.9, t = 42.2, p < 0.001$). Thus, PHRASE POSITION affected DURATION in the last four segments, causing an increase in phrase-final position.

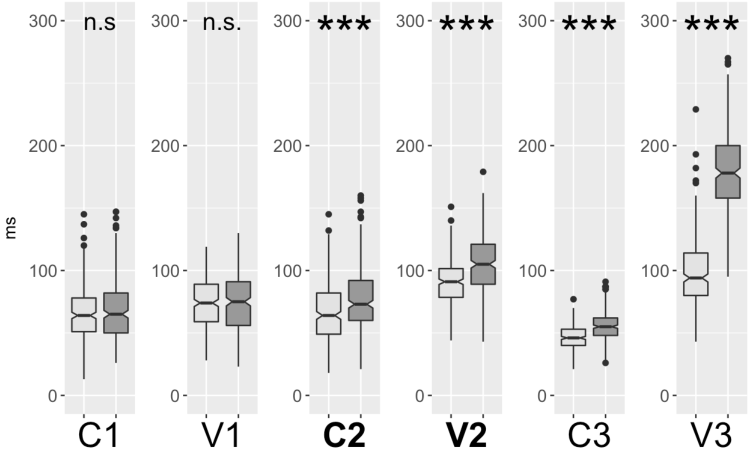


Figure 2: Duration plots (ms) with confidence intervals for the segments of the penultimate stress words with CV.ˈCV.CV structure (e.g., RaMOna) across subjects (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s. not significant; light boxes: phrase-medial, dark boxes: phrase-final)

Figure 3 presents the results for the penultimate stress words with a final coda consonant (e.g., RaMOnas). The post-hoc comparisons did not yield a significant effect for the segments of the antepenultimate syllable (C1: $\beta = 2.2, \mathrm{SE} = 2.3, t = 1, p = 0.323$; V1: $\beta = -0.5, \mathrm{SE} = 2.3, t = -0.2, p = 0.831$). There also was no significant effect for the onset consonant of the penultimate syllable (C2: $\beta = 1.9$, $\mathrm{SE} = 2.3, t = 0.8, p = 0.408$). The plots for the vowel of the penultimate syllable and all following segments clearly suggest longer duration in phrase-final than in phrase-medial position, and the respective comparisons yielded a significant effect (V2: $\beta = 12.9, \mathrm{SE} = 2.3, t = 5.7, p < 0.001$; C3: $\beta = 5.3, \mathrm{SE} = 2.3, t = 2.4$, $p < 0.0183$; V3: $\beta = 65, \mathrm{SE} = 2.3, t = 28.9, p < 0.001$; C4: $\beta = 56.7, \mathrm{SE} = 2.3$, $t = 25.2, p < 0.001$). Thus, phrase position affected duration in the last four segments, causing an increase in phrase-final position.

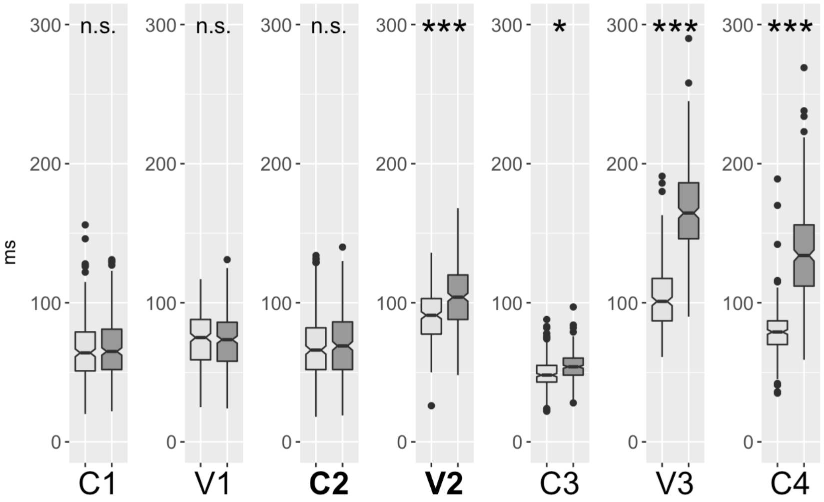


Figure 3: Duration plots (ms) with confidence intervals for the segments of the penultimate stress words with CV.ˈCV.CVC structure (e.g., RaMOnas) across subjects (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s. not significant; light boxes: phrase-medial, dark boxes: phrase-final)

The results for the target words with antepenultimate stress (e.g., KArolin) are illustrated in Figure 4. The coda consonant of the penultimate syllable (C3) was present in only one of the target words (VAlentin) and the coda consonant of the final syllable (C5) was absent in one of the target words (GIsela). The remaining consonants were present in all target words. The plots for the initial onset consonant (C1) suggest a slight tendency towards longer duration in phrase-final position than in phrase-medial position, but the comparison did not yield a significant effect (C1: $\beta = 2.4, \mathrm{SE} = 1.8, t = 1.3, p = 0.187$). The plots for all following segments (V1–C5) clearly suggest longer duration in phrase-final position than in phrase-medial position, and the comparisons yielded a significant effect, respectively (V1: $\beta = 7.6, \mathrm{SE} = 1.8, t = 4.1, p < 0.001$; C2: $\beta = 3.6, \mathrm{SE} = 1.8, t = 2$, $p = 0.0456$; V2: $\beta = 6.7, \mathrm{SE} = 1.8, t = 3.7, p < 0.001$; C3: $\beta = 9.2, \mathrm{SE} = 4.4, t = 2.1$, $p < 0.0359$; C4: $\beta = 6.8, \mathrm{SE} = 1.8, t = 3.7, p < 0.001$; V3: $\beta = 51.8, \mathrm{SE} = 1.8$, $t = 28.4, p < 0.001$; C5: $\beta = 49.9, \mathrm{SE} = 2, t = 25, p < 0.001$). Thus, phrase position affected duration in all three syllables, causing longer duration in phrase-final position than in phrase-medial position.

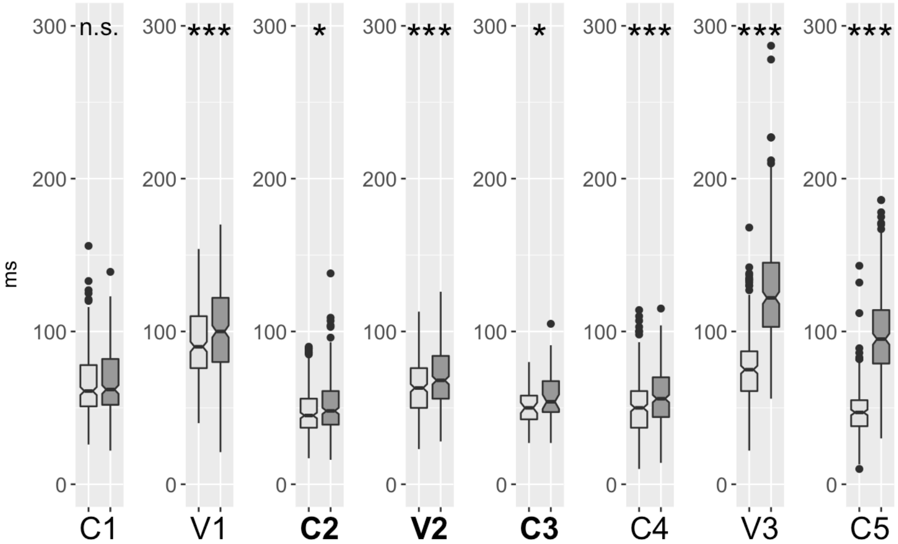


Figure 4: Duration plots (ms) with confidence intervals for the segments of the antepenultimate stress words (e.g., KArolin) across subjects (*** p<.001, ** p< .01, * p<.05, n.s. not significant; light boxes: phrase-medial, dark boxes: phrase-final; C3 was present in only one target word)

3.2 Progressive lengthening

This sub-section presents the results for the distribution of PBL in the target words based on the estimates provided by the LME models. Figure 5 presents the amounts of durational increase in phrase-final compared to phrase-medial position in percentages for the three types of target words. The percentages were calculated as follows: pct = (coefficient value × 100) ÷ intercept value. The dashed line presents the amount of increase for the words with penultimate stress and

CV.ˈCV.CV structure (e.g., RaMOna). The initial onset consonant (C1) involves an increase of 6 percent. The following vowel involves a decrease of 1 percent. After that, on the penultimate syllable, the amount of increase rises to 20 percent on C2 and then slightly falls to 16 percent on V2. On the final syllable, the amount of increase slightly rises to 17 percent on C3 and then undergoes a large increase on the final vowel (V3), reaching 81 percent. The dash-dotted line presents the results for the words with penultimate stress and CV.ˈCV.CVC structure (e.g., RaMOnas). The pattern on the antepenultimate syllable is similar to the prior case. On the penultimate syllable, there is a relatively small increase on C2 (4 percent), followed by a larger increase on V2 (14 percent). On the final syllable, the increase slightly decreases to 12 percent on C3 and then shows a large rise on V3 (63 percent) and further rises on C4 (72 percent). The solid line presents the results for the words with antepenultimate stress (e.g., KArolin). In this case, the increase gradually rises from C1 (5 percent) to C3 (18 percent). After that, there is a large increase on V3 (67 percent) and C4 (104 percent).

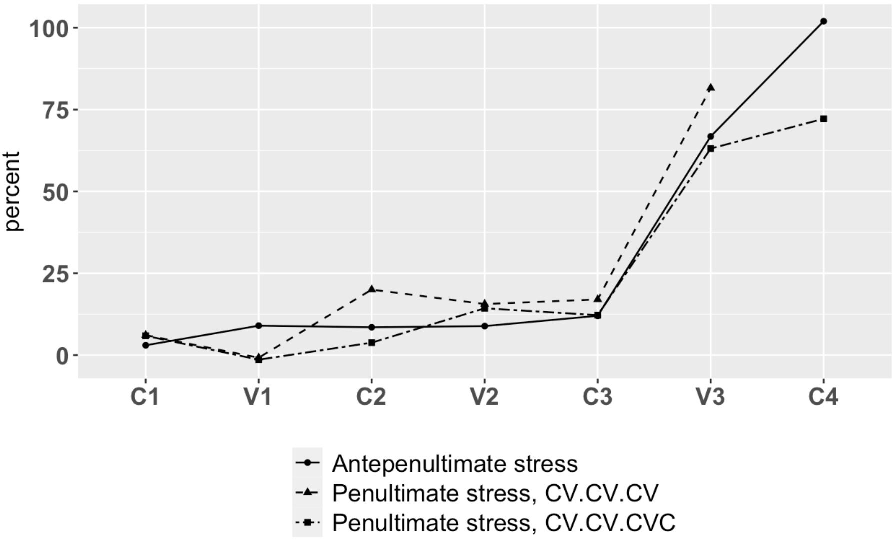


Figure 5: Increase of duration in phrase-final position in percent for the three types of target words across subjects, based on the estimates provided by the LME models (the data for the coda consonant of the penultimate syllable in one of the words with antepenultimate stress is omitted here because the other word forms did not include a coda consonant in this position)

Figure 5 illustrates that the relatively small amount on C1 is comparable across target words. A significant effect was absent in this position across target words. On V1, the words with penultimate stress show an amount of nearly 0 whereas the words with antepenultimate stress show an amount of 9 percent. The comparisons yielded a significant effect only in the latter case. On C2, the types of words with penultimate stress show different patterns: The words with CV.ˈCV.CVC structure involve a considerably smaller increase (4 percent) than the words without such a consonant (20 percent). On the following vowel (V2) and the consonant after that (C3) both types show a similar amount of increase. On the vowel of the final syllable (V3), the types of target words with a following coda consonant show a similar amount of increase (63 and 67 percent, respectively) whereas the target words without such a consonant show a larger amount (82 percent).

It has been found in prior studies that vowels in closed syllables are shorter in duration than vowels in open syllables (e.g., Jones 1950), which is referred to as closed-syllable vowel shortening. Thus, the final vowel in the penultimate stress words might be shorter in duration when a final coda consonant is present than when it is absent. In order to test how much difference in duration between the vowels in these conditions must be attributed to this phenomenon, we compared the duration of the final vowel with and without a final coda consonant in phrase-final position. As shown in Figure 6, the vowels in open syllables were significantly longer than the vowels in closed syllables. A linear mixed effects model accounting for DURATION as a function of CODA CONDITION (levels: absent, present) was fitted to the data of these vowels. Random intercepts and slopes were included for SUBJECT and random intercepts for ITEM. The model estimated that, in closed syllables, the vowel duration was 10 ms shorter than in open syllables ($\beta = 10.2$, SE $= 4.8$, $t = -2.1$). The model was tested against a reduced model without CODA CONDITION as a fixed factor by means of a likelihood ratio test, which yielded a significant effect ($\chi^2(1) = 4.11$, $p = 0.043$). The smaller amount of increase in V3 with a following coda consonant might be related to the fact that the vowel is inherently shorter as well as with the fact that another segment is following (C4, which shows an even larger amount of lengthening). A detailed exploration of a connection between closed syllable shortening and the patterns of PBL is beyond the scope of this study and should be addressed in future research.

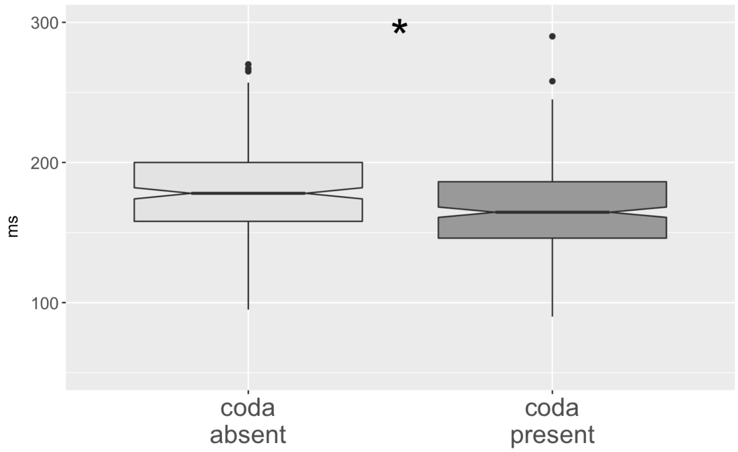


Figure 6: Duration plots (ms) with confidence intervals for the vowel of the final syllable in the penultimate stress words with and without a final coda consonant across subjects (* $p < 0.05$)

4 Discussion and conclusion

4.1 Initiation and scope of PBL

The results suggest that word stress affects the initiation and scope of PBL in German. The initiation of PBL occurred on the main stress syllable across conditions, resulting in a later initiation point in words with penultimate stress than in words with antepenultimate stress. This pattern supports the assumption that the position of main stress serves as an anchor for PBL. The findings, however, deviate from the prediction of the Word Rime hypothesis in one respect: As stated in (1a), this hypothesis predicts that PBL applies to the nuclear vowel of the main stress syllable and all succeeding segments in the phrase-final word (but not to the segments preceding the rime of the main stress syllable). In the penultimate stress words without a final coda consonant (e.g., RaMOna), PBL initiation, however, occurred on the onset consonant of the main stress syllable, that is, the segment immediately preceding the expected initiation point. This suggests a less strict reading of the Word Rime hypothesis. In the other types of target words, the initiation point was on the nuclear vowel of the main stress syllable. The observed connection between PBL initiation and the main stress syllable is compatible with the results by Kohler (1983), who found that PBL occurred across all syllables in di- and trisyllabic words with initial stress. However, the results differ from the observation by Silverman (1990) that the antepenultimate syllable also undergoes PBL in words with penultimate stress in German. Generally, the observed connection between PBL initiation and the main stress syllable is in line with prior findings from other languages (e.g., Berkovits 1994 for Hebrew; White 2002 for British English; Turk & Shattuck-Hufnagel 2007 for American English); yet, there are differences with regard to the details. For example, Turk & Shattuck-Hufnagel (2007) found that PBL is interrupted on the penultimate syllable in words with antepenultimate stress, which was not attested in the present study. It has also been observed that the main stress syllable is relevant for phrase-initial segment lengthening (Napoleão de Souza 2023 [this volume]), which suggests a more general connection between stress positions and boundary-related lengthening.

As for the addition of a coda consonant to the end of the word, the results suggests that the initiation of PBL shifted to the following segment if an additional coda consonant was present. In the penultimate stress words (e.g., RaMOna/ RaMOnas), PBL was initiated on the onset consonant of the penultimate syllable when a final coda consonant was absent, but on the following nuclear vowel when such a consonant was present. Thus, in both conditions, PBL occurred on the portion of the phrase-final word that included the last four segments. Thus, the prediction stated in (1b), capturing the Overlap hypothesis is supported by the present findings. This is in line with prior studies that found an overlap effect (e.g., Byrd & Saltzman 2003 for American English, Seo et al. 2019 for Japanese).

Altogether, our results suggest that both the position of main word stress and the presence/absence of word-final material affect the point of PBL initiation. These findings are compatible with an account that assumes the nuclear vowel of the main stress syllable as the default point of PBL initiation, but allows for PBL on earlier segments if the amount of material between the nuclear vowel and the phrase-boundary is limited. That is, the PBL domain is by default aligned with the nuclear vowel of the main stress syllable at its left and the phrase boundary at its right, but can include earlier segments if this span is too short. This gives rise to the working hypothesis that the PBL domain must have a minimum size in German and thus extends to a segment preceding the nuclear vowel of the main stress syllable. In our data, this occurred when the words had penultimate word stress and lacked a final coda consonant (e.g., RaMOna). The other word forms (e.g., RaMOnas, KArolin) contained enough material between the anchor and the end of the word, so that the PBL domain did not include the preceding onset consonant. This explanation is also compatible with the observation that PBL can occur on material preceding the syllable bearing main word stress, as has been found for words with penultimate stress in German (Silverman 1990) and American English (Cho et al. 2013). The expansion of the PBL domain to earlier material might also be a strategy of the speaker to signal the presence of a prosodic boundary to the listener.[2] This should be addressed in future research.

The conclusions drawn from the comparison between the penultimate stress words with and without a final coda consonant (e.g., RaMOna/RaMOnas) are limited for several reasons in the present study. First, the final coda consonant was always the voiceless alveolar fricative [s]; second, this fricative constitutes a suffix, which yields different morphological structures in the word forms; and, third, the word forms were elicited in different syntactic constructions, which might have affected relative boundary strength. Future research should test mono-morphemic words with various types of coda consonants that are elicited in the same syntactic construction as the words without a final coda consonant.

4.2 Progressive lengthening

We observed a general tendency of progressive lengthening, that is, the amount of PBL gradually increased towards the phrase boundary. This pattern was not consistently applied on the material preceding the final rime, where the amount of PBL showed a slight decrease from one segment to a following one in some cases. This finding suggests that German employs a weak form of progressive lengthening. When the final rime was complex, the amount increased from the nuclear vowel to the following coda consonant, so that the largest amount of PBL always occurred on the final consonant. Furthermore, we found a large increase of PBL on the vowel of the final rime across conditions. That is, the increase of PBL in comparison to the prior segment was strongest on the vowel of the final rime, independent of the rime-internal structure. These patterns are similar to those observed in American English (Turk & Shattuck-Hufnagel 2007) and Japanese (Seo et al. 2019). Unlike to American English (Turk & Shattuck-Hufnagel 2007), PBL was not interrupted on the penultimate syllable in words with antepenultimate stress (e.g., KArolin) in our data.

4.3 PBL in a crosslinguistic perspective

The findings of the present study support the view that the extent of the PBL domain is determined by the position of word stress as well as by the segmental composition of the phrase-final word across languages. Like English (e.g., White 2002, Turk & Shattuck-Hufnagel 2007), German exhibits a connection between PBL initiation and the main stress syllable. Similarly, Greek shows a tendency to pull the initiation of PBL towards the main stress syllable (Katsika 2016). Dutch,

---

[2] Many thanks to an anonymous reviewer for pointing this out to us.

on the other hand, does not show an influence of the main stress syllable, but initiates PBL on the final syllable, unless the final syllable contains only a schwa (Cambier-Langeveld 1997). Altogether, these findings suggest that word stress tends to affect PBL across languages, but languages differ with regard to implementation.

Future research should test the production patterns of PBL with the same or similar materials across languages, as this would provide crosslinguistic data that is directly comparable. The type of materials used in the present study can easily be adapted to other languages. More research is particularly needed on languages with diverse prosodic systems, including languages with an edge-based prosodic system and languages with lexical tone. Schubö et al. (2021) used the same type of materials as in the present study to investigate boundary-related lengthening in Tswana (Southern Bantu), a tone language that expresses specific prosodic boundaries by means of lengthening of the penultimate syllable. They found that PBL occurs on the final syllable in addition to the penultimate lengthening effect, and that the amount of lengthening is comparable on both syllables. This pattern is different from the pattern of progressive lengthening found in German and other languages, which suggests that Tswana has two independent lengthening mechanisms for expressing a prosodic boundary.

The present study found that PBL was strongest on the rime of the final syllable. This suggests that the duration of the final rime might constitute the most salient cue for listeners in the perception of a prosodic boundary based on PBL. Further research is needed on the relevance of the location and amount of lengthening for the perception of a prosodic boundary. Testing the role of these factors for speech perception requires a detailed understanding of their impact on the production of PBL in a given language (see also Turk & Shattuck-Hufnagel 2007 for this point). For example, in order to test if PBL must occur within the designated portion of the phrase-final word or may as well be located on other material near the potential boundary location, we need to understand which factors affect the scope and distribution of PBL. The present study provided insights for German that are essential for such investigations.

Acknowledgements

Many thanks to Nadja Spina for comments and technical support.

Funding information

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project “Preboundary lengthening in a cross-linguistic perspective” (project no. 416902968, GZ ZE 940/3-1), held by Sabine Zerbian.

Appendix A Stimuli

(5) a. Ich werde Ramona oder Karolin und Peter einladen.
b. Ich werde Ramona oder Karolin und Peter einladen.
c. Ich werde Karolin oder Ramona und Peter einladen.
d. Ich werde Karolin oder Ramona und Peter einladen.
e. Ich werde Karolins oder Ramonas und Peters Freunde einladen.
f. Ich werde Karolins oder Ramonas und Peters Freunde einladen.

(6) a. Ich werde Marina oder Salomon und Paula besuchen.
b. Ich werde Marina oder Salomon und Paula besuchen.
c. Ich werde Salomon oder Marina und Paula besuchen.
d. Ich werde Salomon oder Marina und Paula besuchen.
e. Ich werde Salomons oder Marinas und Paulas Oma besuchen.
f. Ich werde Salomons oder Marinas und Paulas Oma besuchen.

(7) a. Ich werde Verena oder Jonathan und Stefan helfen.
b. Ich werde Verena oder Jonathan und Stefan helfen.
c. Ich werde Jonathan oder Verena und Stefan helfen.
d. Ich werde Jonathan oder Verena und Stefan helfen.
e. Ich werde Jonathans oder Verenas und Stefans Schwester helfen.
f. Ich werde Jonathans oder Verenas und Stefans Schwester helfen.

(8) a. Ich werde Rosina oder Valentin und Anna verwarnen.
b. Ich werde Rosina oder Valentin und Anna verwarnen.
c. Ich werde Valentin oder Rosina und Anna verwarnen.
d. Ich werde Valentin oder Rosina und Anna verwarnen.
e. Ich werde Valentins oder Rosinas und Annas Bruder verwarnen.
f. Ich werde Valentins oder Rosinas und Annas Bruder verwarnen.

(9) a. Ich werde Simona oder Fridolin und Lisa suchen.
b. Ich werde Simona oder Fridolin und Lisa suchen.

c. Ich werde Fridolin oder Simona und Lisa suchen.
d. Ich werde Fridolin oder Simona und Lisa suchen.
e. Ich werde Fridolins oder Simonas und Lisas Geschwister suchen.
f. Ich werde Fridolins oder Simonas und Lisas Geschwister suchen.

(10) a. Ich werde Selina oder Gisela und Martin holen.
b. Ich werde Selina oder Gisela und Martin holen.
c. Ich werde Gisela oder Selina und Martin holen.
d. Ich werde Gisela oder Selina und Martin holen.
e. Ich werde Giselas oder Selinas und Martins Eltern holen.
f. Ich werde Giselas oder Selinas und Martins Eltern holen.

(11) a. Ich werde Ramona oder Salomon und Anna befragen.
b. Ich werde Ramona oder Salomon und Anna befragen.
c. Ich werde Salomon oder Ramona und Anna befragen.
d. Ich werde Salomon oder Ramona und Anna befragen.
e. Ich werde Salomons oder Ramonas und Annas Bruder befragen.
f. Ich werde Salomons oder Ramonas und Annas Bruder befragen.

(12) a. Ich werde Marina oder Jonathan und Stefan abholen.
b. Ich werde Marina oder Jonathan und Stefan abholen.
c. Ich werde Jonathan oder Marina und Stefan abholen.
d. Ich werde Jonathan oder Marina und Stefan abholen.
e. Ich werde Jonathans oder Marinas und Stefans Geschwister abholen.
f. Ich werde Jonathans oder Marinas und Stefans Geschwister abholen.

(13) a. Ich werde Rosina oder Valentin und Lisa beschuldigen.
b. Ich werde Rosina oder Valentin und Lisa beschuldigen.
c. Ich werde Valentin oder Rosina und Lisa beschuldigen.
d. Ich werde Valentin oder Rosina und Lisa beschuldigen.
e. Ich werde Valentins oder Rosinas und Lisas Freunde beschuldigen.
f. Ich werde Valentins oder Rosinas und Lisas Freunde beschuldigen.

(14) a. Ich werde Verena oder Fridolin und Martin anrufen.
b. Ich werde Verena oder Fridolin und Martin anrufen.
c. Ich werde Fridolin oder Verena und Martin anrufen.
d. Ich werde Fridolin oder Verena und Martin anrufen.

e. Ich werde Fridolins oder Verenas und Martins Eltern anrufen.
f. Ich werde Fridolins oder Verenas und Martins Eltern anrufen.

(15) a. Ich werde Simona oder Gisela und Peter begleiten.
b. Ich werde Simona oder Gisela und Peter begleiten.
c. Ich werde Gisela oder Simona und Peter begleiten.
d. Ich werde Gisela oder Simona und Peter begleiten.
e. Ich werde Giselas oder Simonas und Peters Oma begleiten.
f. Ich werde Giselas oder Simonas und Peters Oma begleiten.

(16) a. Ich werde Selina oder Karolin und Paula ermahnen.
b. Ich werde Selina oder Karolin und Paula ermahnen.
c. Ich werde Karolin oder Selina und Paula ermahnen.
d. Ich werde Karolin oder Selina und Paula ermahnen.
e. Ich werde Karolins oder Selinas und Paulas Schwester ermahnen.
f. Ich werde Karolins oder Selinas und Paulas Schwester ermahnen.

Appendix B LME model outputs

Table 4: LME model output for the words with penultimate stress and CV.ˈCV.CV structure (significant *t* values are boldfaced)

	Estimate	SE	*t*
(Intercept)	98.4	1.8	**55.4**
phrase position *final*	80.1	1.9	**42.2**
segment position *2*	51.4	1.8	**−28.2**
segment position *3*	−7.7	1.8	**−4.2**
segment position *4*	−32.1	1.8	**−17.6**
segment position *5*	−24.7	1.8	**−13.6**
segment position *6*	−32.7	1.8	**−18.0**
phrase position *final* : segment position *2*	−72	2.6	**−27.8**
phrase position *final* : segment position *3*	−66.2	2.6	**−25.6**
phrase position *final* : segment position *4*	−69.4	2.6	**−26.9**
phrase position *final* : segment position *5*	−81	2.6	**−31.3**
phrase position *final* : segment position *6*	−76.5	2.6	**−29.6**

Table 5: LME model output for the words with penultimate stress and CV.ˈCV.CVC structure (significant *t* values are boldfaced)

	Estimate	SE	*t*
(Intercept)	79	1.8	**43.3**
phrase position *final*	56.7	2.3	**25.2**
segment position *2*	24.3	1.9	**12.9**
segment position *3*	−29.9	1.9	**−15.9**
segment position *4*	12.2	1.9	**6.5**
segment position *5*	−10.4	1.9	**−5.6**
segment position *6*	−5.3	1.9	**−2.8**
segment position *7*	−13.2	1.9	**−7.0**
phrase position *final* : segment position *2*	8.3	2.6	**3.2**
phrase position *final* : segment position *3*	−51.4	2.6	**−20.1**
phrase position *final* : segment position *4*	−43.8	2.6	**−17.1**
phrase position *final* : segment position *5*	−54.8	2.6	**−21.4**
phrase position *final* : segment position *6*	−57.2	2.6	**−22.4**
phrase position *final* : segment position *7*	−54.5	2.6	**−21.3**

Table 6: LME model output for the words with penultimate stress and CV.ˈCV.CVC structure (significant *t* values are boldfaced)

	Estimate	SE	*t*
(Intercept)	50.8	3.6	**14.2**
phrase position *final*	50	2.0	**25.0**
segment position *2*	24	2.0	**12.3**
segment position *3*	−0.3	2.0	−0.4
segment position *4*	10.1	3.4	**3.0**
segment position *5*	12.8	2.0	**6.6**
segment position *6*	−3.3	2.0	**−1.7**
segment position *7*	42.1	2.0	**21.6**
segment position *8*	14.3	2.0	**7.3**
phrase position *final* : segment position *2*	1.9	2.7	0.7
phrase position *final* : segment position *3*	−43.1	2.7	**−16.0**
phrase position *final* : segment position *4*	−40.7	4.8	**−8.5**
phrase position *final* : segment position *5*	−43.2	2.7	**−16.0**
phrase position *final* : segment position *6*	−46.3	2.7	**−17.1**
phrase position *final* : segment position *7*	−42.3	2.7	**−15.7**
phrase position *final* : segment position *8*	−47.5	2.7	**−17.6**

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: 10.18637/jss.v067.i01.

Berkovits, Rochele. 1993a. Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech* 36(1). 89–98. DOI: 10.1177/002383099303600105.

Berkovits, Rochele. 1993b. Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics* 21. 479–489. DOI: 10.1016/S0095-4470(19)30231-1.

Berkovits, Rochele. 1994. Durational effects in final lengthening, gapping and contrastive stress. *Language and Speech* 37. 237–250. DOI: 10.1177/002383099403700302.

Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer*. http://www.praat.org/.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155–180. DOI: 10.1159/000261913.

Byrd, Dani, Jelena Krivokapić & Sungbok Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America* 120. 1589–1599. DOI: 10.1121/1.2217135.

Byrd, Dani, Sungbok Lee, Daylen Riggs & Jason Adams. 2005. Interacting effects of syllable and phrase position on consonant articulation. *The Journal of the Acoustical Society of America* 118. 3360–3873. DOI: 10.1121/1.2130950.

Byrd, Dani & Daylen Riggs. 2008. Locality interactions with prominence in determining the scope of phrasal lengthening. *Journal of the International Phonetic Association* 38(2). 187–202. DOI: 10.1017/s0025100308003460.

Byrd, Dani & Elliot Saltzman. 2003. The elastic phrase: Modelling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31. 149–180. DOI: 10.1016/s0095-4470(02)00085-2.

Cambier-Langeveld, Tina. 1997. The domain of final lengthening in the production of Dutch. In Jane A. Coerts & Helen de Hoop (eds.), *Linguistics in the Netherlands*, 13–24. Amsterdam: Benjamins. DOI: 10.1075/avt.14.04cam.

Cambier-Langeveld, Tina. 2000. *Temporal marking of accents and boundaries*. University of Amsterdam. (Doctoral dissertation). https://hdl.handle.net/11245/1.172937.

Cho, Taehong, Jiseung Kim & Sahyang Kim. 2013. Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *The*

Journal of the Acoustical Society of America 133. 384–390. DOI: 10.1121/1.4800179.

Delattre, Pierre. 1965. *Comparing the phonetic features of English, French, German and Spanish: An interim report*. Heidelberg: Julius Groos.

Dogil, Grzegorz & Briony Williams. 1999. The phonetic manifestation of word stress. In Harry van der Hulst (ed.), *Word prosodic systems in the languages of Europe*, 273–334. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110197082.1.273.

Féry, Caroline & Frank Kügler. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36. 680–703. DOI: 10.1016/j.wocn.2008.05.001.

Goldstein, Louis, Dani Byrd & Elliot Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.), *Action to language via the mirror neuron system*, 215–249. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511541599.008.

Grice, Martine, Stefan Baumann & Ralf Benzmüller. 2005. German intonation in autosegmental-metrical phonology. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 55–83. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.003.0003.

Hall, Tracy Alan. 1992. *Syllable structure and syllable-related processes in German*. Tübingen: Niemeyer. DOI: 10.1515/9783111657882.

Hofhuis, Elise, Carlos Gussenhoven & Toni Rietveld. 1995. Final lengthening at prosodic boundaries in Dutch. In *Proceedings of the 13th ICPhS*, 14–19. http://hdl.handle.net/2066/105391.

Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3). 346–363. DOI: 10.1002/bimj.200810425.

Huttenlauch, Clara, Carola de Beer, Sandra Hanne & Isabell Wartenburger. 2021. Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Laboratory Phonology* 12(1). 1–31. DOI: 10.5334/labphon.221.

Huttenlauch, Clara, Marie Hansen, Carola de Beer, Sandra Hanne & Isabell Wartenburger. 2023. Age effects on linguistic prosody in coordinates produced to varying interlocutors: Comparison of younger and older speakers. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 157–192. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777534.

Jones, Daniel. 1950. *The phoneme: Its nature and use*. Cambridge: Heffer. DOI: 10.2307/410108.

Katsika, Argyro. 2016. The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics* 55. 149–181. DOI: 10.1016/j.wocn.2015.12.003.

Kentner, Gerrit & Caroline Féry. 2013. A new approach to prosodic grouping. *The Linguistic Review* 30. 277–311. DOI: 10.1515/tlr-2013-0009.

Klatt, Dennis H. 1975. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3. 129–140. DOI: 10.1016/s0095-4470(19)31360-9.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59. 1208–1220. DOI: 10.1121/1.380986.

Kohler, Klaus J. 1983. Prosodic boundary signals in German. *Phonetica* 40(2). 445–472. DOI: 10.1159/000261685.

Krull, Diana. 1997. Prepausal lengthening in Estonian: Evidence from conversational speech. In Ilse Lehiste & Jaan Ross (eds.), *Estonian prosody: Papers from a symposium*, 136–148.

Nakai, Satsuki, Sari Kunnari, Alice E. Turk, Kari Suomi & Riikka Ylitalo. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37. 29–45. DOI: 10.1016/j.wocn.2008.08.002.

Napoleão de Souza, Ricardo. 2023. Segmental cues to IP-initial boundaries: Data from English, Spanish, and Portuguese. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 35–86. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777528.

Peters, Benno, Klaus J. Kohler & Thomas Wesener. 2005. Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In Klaus J. Kohler, Felicitas Kleber & Benno Peters (eds.), *Prosodic structures in German spontaneous speech*, 143–184. Kiel: Arbeitsberichte (AIPUK). https://www.ipds.uni-kiel.de/kjk/pub_exx/aipuk35a/aipuk35a_6.pdf.

Peters, Jörg. 2018. Phonological and semantic aspects of German intonation. *Linguistik online 88*. 88–107. DOI: 10.13092/lo.88.4191.

Petrone, Caterina, Hubert Truckenbrodt, Caroline Wellmann, Julia Holzgrefe-Lang, Isabell Wartenburger & Barbara Höhle. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61. 71–92. DOI: 10.1016/j.wocn.2017.01.002.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Schubö, Fabian, Ian Bekker, Rigardt Pretorius, Valecia Wagner & Sabine Zerbian. 2021. An exploration into penultimate and final lengthening in Tswana (Southern Bantu). *Stellenbosch Papers in Linguistics Plus* 62. 17–37. DOI: 10.5842/62-0-891.

Schubö, Fabian, Anna Roth, Viviana Haase & Caroline Féry. 2015. Experimental investigations on the prosodic realization of restrictive and appositive relative clauses in German. *Lingua* 154. 65–86. DOI: 10.1016/j.lingua.2014.11.006.

Schubö, Fabian & Sabine Zerbian. 2020. Phonetic content and phonological structure affect pre-boundary lengthening in German. In *Proceeding of Speech Prosody*, 111–115. DOI: 10.21437/speechprosody.2020-23.

Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Mark Aronoff & Richard T. Oehrle (eds.), *Language sound structure: Studies in phonology*, 107–136. Cambridge, MA: MIT Press.

Seo, Jungyun, Sahyang Kim, Haruo Kubozono & Taehong Cho. 2019. Interaction between rhythmic structure and preboundary lengthening in Japanese. In *Proceedings of the 19th ICPhS*, 637–641. https://www.assta.org/proceedings/ICPhS2019/papers/ICPhS_686.pdf.

Silverman, Kim E. A. 1990. The separation of prosodies; comments on Kohler's paper. In John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech*, 521–602. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511627736.008.

Snedeker, Jesse & John Trueswell. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48. 103–130. DOI: 10.1016/s0749-596x(02)00519-3.

Truckenbrodt, Hubert. 2002. Upstep and embedded register levels. *Phonology* 19. 77–120. DOI: 10.1017/s095267570200427x.

Truckenbrodt, Hubert. 2006. Phrasal stress. In Keith Brown (ed.), *The encyclopedia of languages and linguistics*, vol. 9, 572–579. Oxford: Elsevier. DOI: 10.1016/b0-08-044854-2/04447-3.

Truckenbrodt, Hubert. 2007. Upstep on edge tones and on nuclear accents. In Carlos Gussenhoven & Tomas Riad (eds.), *Tones and tunes: Experimental studies in word and sentence prosody*, 349–386. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110207576.2.349.

Turk, Alice E., Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. In Stefan Sudhoff, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.), *Methods in empirical prosody research*, 1–28. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110914641.1.

Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35(4). 445–472. DOI: 10.1016/j.wocn.2006.12.001.

Vaissière, Jacqueline. 1983. Language-independent prosodic features. In Anne Cutler & D. Robert Ladd (eds.), *Prosody: Models and measurements*, 53–66. Berlin: Springer. DOI: 10.1007/978-3-642-69103-4_5.

Wagner, Michael. 2005. *Prosody and recursion*. Massachusetts Institute of Technology. (Doctoral dissertation). https://dspace.mit.edu/handle/1721.1/33713.

Wellmann, Caroline, Julia Holzgrefe-Lang, Hubert Truckenbrodt, Isabell Wartenburger & Barbara Höhle. 2023. Developmental changes in prosodic boundary cue perception in German-learning infants. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 119–156. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777532.

White, Laurence. 2002. *English speech timing: A domain and locus approach*. University of Edinburgh. (Doctoral dissertation). https://www.cstr.ed.ac.uk/projects/eustace/fulltext/front.pdf.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Chapter 2

Segmental cues to IP-initial boundaries: Data from English, Spanish, and Portuguese

Ricardo Napoleão de Souza
University of Edinburgh

This study investigates the segmental phonetic correlates of IP-initial boundaries in unstressed syllables in three lexical stress languages: English, Spanish, and Portuguese. Using acoustic data gathered under similar experimental conditions, it tests the hypothesis that domain-initial strengthening cues prosodic boundaries in language-specific ways. Moreover, it investigates the role of lexical stress in the phenomenon by focusing on unstressed syllables in post-boundary position, while at the same time testing whether the scope of the strengthening is indeed restricted by measuring segments away from the putative boundary. Results from the analyses of 14 speakers of each of the languages ($N = 42$) strengthen the case for language-specific effects; however, the data suggest that segments farther away from the IP boundary show strengthening as well.

1 Introduction

In written language, capitalization signals the beginning of a new sentence, and punctuation marks its end. In speech, various phonetic adjustments play a role similar to capitalization and punctuation, in that they mark the beginnings and endings of prosodic units. In addition to its suprasegmental dimension, a growing body of experimental investigations suggests that prosodic structure influences the phonetic realization of segments in systematic ways (see Fougeron 2001, Cho 2015, for detailed reviews). What's more, the results of these studies suggest that

Ricardo Napoleão de Souza. 2023. Segmental cues to IP-initial boundaries: Data from English, Spanish, and Portuguese. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 35–86. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777528

speakers use language-specific, gradient phonetic detail to group their speech into meaningful units (e.g. Keating et al. 2003, Cho & McQueen 2005).

The investigation of the ways in which prosodic structure modulates phonetic information thus offers important contributions to our understanding of how speech is organized (see also Shattuck-Hufnagel & Turk 1996, Fletcher 2010). Looking at the edges of prosodic units allows speech scientists to determine how the flow of information present in discourse is parsed into cognitively manageable units (Krivokapić 2014). While there is a great deal of research on prosodic domain endings (see Cole 2015 for a comprehensive review), the phonetic encoding of the beginning of prosodic units is much less understood.

There are four main unresolved issues regarding the phonetic manifestation of domain-initial boundaries. First, it is still unclear how boundary-initial prominence relates to other levels of prominence, such as lexical stress (cf. Turk & Shattuck-Hufnagel 2007). Secondly, the evidence on the scope of its temporal effect with regard to the prosodic edge is inconclusive (for a discussion, see Katsika 2016). Thirdly, research has yet to determine the role of language-specific phonology in the phonetic manifestation of domain-initial effects (cf. Cho 2016). Lastly, because most of the literature on domain beginnings has focused on articulation, data on how initial boundaries translate into the *acoustic* signal are still relatively scarce (see Section 1.1 below).

This chapter aims to address these four issues by investigating the acoustic effects of domain-initial IP boundaries on unstressed syllables in prenuclear position in English, Spanish, and Portuguese. The examination of unstressed syllables has the benefit of isolating domain-initial effects from lexical stress, while also allowing for an assessment of their scope of influence. Moreover, a comparison of languages whose unstressed syllables show distinct phonetic properties through similar experimental materials allows for a more straightforward assessment of how language-specific word prosody patterns influence the phenomenon. Since the acoustic results presented here are based on the analysis of 42 speakers of American English, Mexican Spanish and Brazilian Portuguese, results provide more robust quantitative data that may disentangle some of the issues in previous small-scale articulatory studies.

1.1 Prosodic boundaries and their markings

One of the central goals of the study of prosody is understanding how speakers group chunks of speech into coherent units according to their meaning and pragmatic functions (cf. Selkirk 1984, Nespor & Vogel 1986, Jun 2006). Although the number of prosodic domains postulated for each language varies, each of these units is separated from each other through phonetically cued prosodic boundaries (Nespor & Vogel 1986, Shattuck-Hufnagel & Turk 1996, among others). Acoustic correlates of prosodic boundaries include tonal markings, higher intensity and F0 changes, and adjustments in duration (Fletcher 2010). These may differ between initial and final boundaries.

Words that immediately precede a prosodic phrase boundary (i.e. the final edge) have been shown to be consistently longer than those occurring elsewhere in the phrase (e.g. Jun & Beckman 1994, Wightman et al. 1992, Gussenhoven & Rietveld 1992, Berkovits 1994, Byrd et al. 2006, Fougeron 2001, Turk & Shattuck-Hufnagel 2000, Frota & Prieto 2015, among many others). Referred to as PHRASE-FINAL LENGTHENING, this type of boundary marking effect has been detected most consistently in the very last syllable of the phrase-final word (cf. Turk & Shattuck-Hufnagel 2007).

On the other hand, boundary marking on the segment near the beginning of prosodic domain has received much less attention in the literature. There is evidence, however, that segments that immediately follow a prosodic boundary (e.g. at the initial edge of an IP) are produced with stronger articulation than when they occur elsewhere in the phrase (cf. Dilley et al. 1996, Fougeron & Keating 1997, Fougeron 1999, Cho & Keating 2001, 2009, Keating et al. 2003, Cho & McQueen 2005, Georgeton & Fougeron 2014, Cho 2011, among others). Because these effects have been noted first in the articulatory realm, boundary-initial marking is most often referred to as DOMAIN-INITIAL STRENGTHENING in the literature. Importantly, the articulatory evidence suggests that domain-initial edges are marked solely on the boundary-adjacent segment.

Domain-initial strengthening has been reported in several languages (Cho 2015), although the levels at which it is significant vary depending on the language (e.g. Keating et al. 2003), as well as on individual studies. More specifically, there are reports of segmental correlates of boundary-initial prominence in two of the three languages in the current sample: English (Pierrehumbert & Talkin 1992, Fougeron & Keating 1997, Byrd et al. 2006, inter alia) and Spanish (Lavoie 2001, Parrell 2014). However, neither Lavoie (2001) nor Parrell (2014) set out to investigate boundary prominence in Spanish, even though both describe results that could be interpreted as domain-initial strengthening. To the best of my knowledge, no investigation has examined domain-initial strengthening in Portuguese hitherto.

ACOUSTIC EVIDENCE for domain-initial strengthening, on the other hand, is much more limited given that it is often reported secondarily in articulatory studies (e.g. Katsika 2016 for Greek; Cho & McQueen 2005 for Dutch; Oller 1973, Tabain 2003 for English; Lavoie 2001, Parrell 2014 for Spanish, Hsu & Jun 1998 for

Taiwanese and Korean). Studies that specifically investigated the acoustic correlates of domain-initial strengthening include Cho et al. (2007), Cole et al. (2007), and White et al. (2020) for English, Italian and Hungarian; Kuzla et al. (2007) for German; and Bodur et al. (2021) for Turkish. Some of the acoustic correlates of domain-initial marking for consonants include voice onset time (VOT), the occurrence of stop burst releases, closure duration, degrees of voicing, duration of nasal murmur, among others. Because individual studies typically investigate one or two variables at a time, it is yet to be determined how different languages use most of those acoustic properties to cue initial edges. Importantly, most of the studies above looked at syllables that were either lexically or phrasally stressed (e.g. under prosodic focus) so that at least some of the effects observed may derive from other types of prosodic elements.

Still, VOT is perhaps the most commonly reported variable connected to domain-initial strengthening. Differences in VOT values correlate with boundary markings on consonants in English (e.g. Pierrehumbert & Talkin 1992, Beckman et al. 1992, Cho & Keating 2001), Dutch (Cho et al. 2007), Korean (Cho & Keating 2001, among others), all of which use VOT to distinguish between /p t k/ and /b d g/ (see Cho et al. 2019 for a discussion). However, VOT is also reported as a cue to initial boundaries in languages with short-lag VOT consonants, such as French (Fougeron 2001) or Japanese (Onaka et al. 2003). The only acoustic correlate of boundary marking on voiceless stops in Spanish is the occurrence of stop burst releases (Lavoie 2001, also for English).

As with the articulatory data, these studies suggest a local effect of domain-initial strengthening, meaning that segmental change due to proximity to the domain-initial boundary occurs on the very first segment following the prosodic edge. On the other hand, studies investigating nasals in nasal consonant-vowel sequences at the IP boundary have found that the vowel in those syllables shows less nasalization (e.g. Cho et al. 2017), which hints at a larger scope of actuation than the initial segment. However, many of the studies reviewed here only investigated individual segments, leaving the question of the scope of the effect somewhat open (see Section 1.3 below).

The body of work presented above suggests that the phonetic variation associated with domain-initial positions correlates systematically with various phrase levels within language-specific prosodic structure. However, a cohesive account of the phenomenon is still lacking despite a growing interest in the so-called prosody-phonetics interface. There are three possible explanations to this observation. First, the number of specificities linked to domain-initial effects reduces comparability between different studies of the phenomenon. Secondly, the interaction of boundary markings with different kinds of prosodic prominence has often been overlooked in previous research, thereby introducing important confounds. Thirdly, methodological choices in previous studies, including the choice of materials and sample sizes, introduce non-trivial challenges in the interpretation of results. In order to address these issues, the current study uses the same methodology to investigate domain-initial effects in three lexical stress languages: English, Spanish, and Portuguese. The next section addresses key aspects of prosodic structure in each of the three languages.

1.2 Prosodic aspects of English, Spanish, and Portuguese

1.2.1 Prosodic structure and phrasal prominence

The grouping function of prosody is assumed here to follow a hierarchical structure, meaning that higher levels of structure contain the lower levels, both of which are language-specific (Jun 2014). At the same time, the two highest levels in the prosodic structure, namely the Utterance and the Intonational Phrase (IP), are among the most frequent across languages (cf. Jun 2006, 2014). The IP has been identified specifically as the major prosodic level around which the phonetic correlates of domain-initial boundaries can be measured (cf. Keating et al. 2003).

Indeed, the literature describes the IP as a major domain in English, Spanish, and Portuguese alike (e.g. Beckman & Pierrehumbert 1986, Prieto et al. 1995, Frota 2000). Since there is less agreement for levels below the IP in the three languages, especially regarding mid-level domains (cf. Shattuck-Hufnagel & Turk 1996, Frota & Prieto 2015), this study makes no specific claims regarding levels other than the IP.

English, Spanish, and Portuguese also share other prosodic features. In the three languages, final IP boundaries in declarative utterances are associated with acoustic markings such as final lengthening, pitch declines, and pauses (Beckman & Pierrehumbert 1986, Prieto et al. 1995, Frota 2000). Additionally, the locus of the nuclear accent is similar in the three languages. In neutral declarative sentences, the nuclear accent tends to fall on the rightmost lexical word of the IP and is anchored on the stressed syllable of that word (Pierrehumbert 1980, Hualde & Prieto 2015, Frota & Moraes 2016). Moreover, the nuclear accent can be moved around within the phrase in English, Spanish, and Portuguese, so that any word can potentially receive emphasis (Beckman & Edwards 1994, Ladd 2008, Frota & Prieto 2015, Frota & Moraes 2016; see also Vogel et al. 2018). This prosodic feature common to the three languages was useful in the design of the stimuli for the reading task, as explained in Section 2.2.

On the other hand, the distribution of prenuclear pitch accents differs in English, Spanish, and Portuguese. Spanish and (Brazilian) Portuguese are described as languages with a dense distribution of pitch accents in non-question intonation, so that "essentially every prosodic word (…) receives a pitch accent" (Frota & Prieto 2015: 397).[1] In English, pitch accents other than the nuclear accent are less common than in Romance as a whole, though speakers may accent prenuclear elements (Ladd 2008). Factors influencing the placement of pitch accents in English include semantic-pragmatic factors, structural factors, and rhythmic factors (Shattuck-Hufnagel & Turk 1996).

1.2.2 Lexical stress

As hinted above, English, Spanish, and Portuguese all have lexical stress. Lexical stress is obligatory for content words in these languages (Liberman & Prince 1977, Hualde 2012, Mateus & D'Andrade 2000). Lexical stress is also contrastive in all three (e.g. English [ˈpʰɜɻmɪt] 'a permit' vs. [pʰəɻˈmɪt] 'to permit'; Spanish [ˈnumeɾo] 'number' vs. [nuˈmeɾo] 'I number' vs. [numeˈɾo] 'she numbered'; Portuguese [ˈmɛdʒɪkʊ] 'a doctor' vs. [meˈdʒikʊ] 'I medicate' vs. [medʒɪˈko] 'she medicated').

Stress placement is considered free and difficult to predict in English (e.g. Liberman & Prince 1977). In Spanish and Portuguese, stress follows somewhat more regular patterns. In polysyllabic words, lexical stress generally falls on any one of the three last syllables (Hualde 2012, Mateus & D'Andrade 2000). Although variable, the placement of lexical stress in Spanish and Portuguese can usually be determined based on a series of morphosyntactic and phonological patterns. The main acoustic correlates of lexical stress in the three languages are duration, F0 anchoring, and amplitude (Beckman 1986, Hualde 2012, Mateus & D'Andrade 2000).

Despite the similarities described above, the three languages differ substantially in the degrees to which lexical stress impacts unstressed syllables. Fully unstressed syllables (i.e. not bearing secondary stress) are much shorter than their stressed counterparts in both English and Portuguese (cf. Plag et al. 2011, Cantoni 2013), whereas that difference is less pronounced in Spanish (e.g. Ortega-Llebaria & Prieto 2007). Unstressed consonants show phonetic differences in both English and Spanish, but not in Brazilian Portuguese (Cristófaro-Silva et al. 2019).

---

[1] A prosodic word in Spanish or Portuguese can be defined as a lexical word plus any unstressed clitics (Hualde 2007, Vigário 2003). For instance, the Spanish article *el* in neutral statements such as *el dinosaurio* /eldinoˈsauɾjo/ 'the dinossaur'; or the Portuguese *se* in *feriu-se* /fɪˈɾiʊsɪ/ 'she hurt herself'. Test words in the current study never contained such clitics and are thus classified as lexical words.

Fully unstressed vowels in English are often centralized to [ə], whereas none of the five Spanish vowels /i e a o u/ shows much qualitative change when unstressed (Nadeu 2014). In Portuguese, fully unstressed vowels show several patterns of reduction, depending on vowel quality, nasality, distance from the stressed syllable and position within the word (Crosswhite 2001, Mateus & D'Andrade 2000, Cristófaro-Silva et al. 2019). For instance, while Câmara JR (1972) described that only five of the seven oral vowels may occur in prestressed position in Brazilian Portuguese (i.e. /i e a o u/ out of /i e ɛ a ɔ o u/), further variable reduction is now common in those unstressed syllables. Prestressed /i e/ may occur as [ɨ ɪ ɪ̥ ʲ ∅], while prestressed /o u/ may appear as [ʊ ʊ̥ ʷ ∅]. In short, English shows the most consolidated patterns of vowel reduction, whereas Portuguese has been undergoing a series of changes that seem to relate to lexical stress. Spanish unstressed vowels, on the other hand, remain largely unaffected by stress-related reductions. The next section describes how still unresolved issues concerning domain-initial strengthening can be elucidated through an investigation of the languages in the present sample, as well as specific hypotheses and predictions regarding their behavior following IP-initial edges.

1.3 Unresolved issues regarding domain-initial strengthening

The three main unresolved issues regarding domain-initial strengthening are their relationship to lexical stress (and other types of prosodic prominence), their scope of influence, and the specific ways in which they interact with language-specific segmental phonology. Another important caveat lies in the fact that most research on domain-initial strengthening has focused on articulatory data from a few speakers. It is possible that inconclusive results regarding those issues could derive in part from experimental design choices and/or from small sample sizes.

While some studies manipulate nuclear accent (e.g. Cho et al. 2017), others may have inadvertently introduced focus accents by having speakers repeat similar or identical carrier sentences (e.g. Lavoie 2001, or Parrell 2014 for Spanish) in which only the test words vary. As a result, these carrier sentences would have likely elicited contrastive focus accents on test words (see Roettger & Gordon 2017 for a discussion). Additionally, a lack of control for other levels of prosodic prominence in test words creates difficulties for the interpretation of experimental findings (cf. Fougeron 2001: 112).

Due to the challenges of collecting articulatory data, many studies have been conducted using small samples of three to five speakers per study. The understandably limited number of speakers in these studies may have nonetheless allowed idiosyncrasies in the speech materials or the behavior of participants to influence the results. In smaller samples, individual differences in speaker behavior may skew results in ways that make it difficult to distinguish the effects of the variables being tested from those relating to idiosyncrasies in the speech of participants.

More specifically, the current study seeks to provide answers to three research questions (RQ 1–3) stated below.

RQ1: How is domain-initial strengthening expressed acoustically?

While predominantly articulatory in nature, studies of domain-initial strengthening have identified a number of acoustic properties that showed an impact of position within the prosodic domain. Generally, domain-initial strengthening has been found to increase the saliency of segments following the prosodic boundary (Cho 2016), so it is hypothesized that acoustic properties of boundary-adjacent segments will also show an increase in their magnitude. Specifically, the current study is designed to evaluate VOT, burst releases at stop closure, vowel duration, F1 and vowel dispersion.

VOT is the acoustic property that has most often been associated with domain-initial strengthening in several languages (Cho 2016, see also Section 1.1 above). For /p t k/, it is hypothesized that VOT will show greater values following the IP boundary than phrase-medially. Additionally, Lavoie (2001) found differences in the occurrence of stop release bursts in consonants at the beginning versus in the middle of words in both English and Spanish. Consonant bursts are associated with articulatory strengthening (cf. Stevens & Keyser 1989, Torreira & Ernestus 2011). Expanding on the pattern in Lavoie (2001), it is hypothesized that in the current investigation, stops following an IP boundary will show burst releases more often than those occurring IP-medially. Put differently, distance from the prosodic boundary is expected to correlate with more burstless stop releases.

For vowels, Cho & Keating (2009) found partial evidence that domain-initial strengthening increases first formant values in vowels in CV syllables following an IP boundary (see Oh 2021 for similar results for Brazilian Portuguese). Additionally, first formant values serve as an indirect measure of jaw opening, which has been found to correlate with prosodic properties such as prosodic focus (e.g. Erickson 1998). It is expected that vowels showing domain-initial strengthening will thus show higher F1 values than those occurring mid-phrase.

Data on vowel duration is less conclusive, on the other hand. Whereas Oh (2021) found effects of prosodic position within the word (i.e. word-initial vowels in unstressed CV syllables were longer) for Portuguese, target words in her study may have been phrasally accented. Cho & Keating (2009) report no effect on duration based on proximity to the domain-initial edge, but target syllables in their study received either primary or secondary stress. Although data on vowels in post-boundary CV syllables suggest little to no effect of domain-initial strengthening, potential confounds with lexical and phrasal prominence warrant further testing of its scope (see discussion in Cho 2016).

If the effects of domain-initial strengthening are indeed restricted to the edge-adjacent segment, the vowels in test CV syllables will show no durational differences between prosodic contexts. Alternatively, the hypothesized longer VOT at the IP-initial boundary might increase the overlap between consonant and vowel gestures, leading to shorter and/or devoiced vowels. However, longer VOT at IP-initial CV syllables would still primarily refer to the first segment following the boundary, with any possibly effects on the second segment resulting from assimilation to the former.

RQ2: To what extent is the acoustic manifestation of domain-initial strengthening manifested language-specifically?

Cho & McQueen (2005) and Cho (2016) discuss how different languages show specific combinations of acoustic cues to boundary marking. According to Cho, segments subject to domain-initial strengthening are "fine-tuned (…) making reference to the phonetic content provided by the language-specific phonetic feature system" (Cho 2016: 136). In other words, the phonetic expression of initial boundary marking depends on which features the language already uses to convey phonological distinctions such as /k/ vs. /g/, or /ʊ/ vs. /u/. The current study uses VOT and differences in unstressed vowel qualities to test Cho's hypothesis.

For instance, VOT serves different roles in English vs. Spanish or Portuguese. In English, /p t k/ are distinguished from /b d g/ mostly through VOT, whereas Spanish and Portuguese primarily use voicing to signal the same distinction. Recent data on Brazilian Portuguese, however, suggest that longer VOT ("aspiration") may be emerging as an acoustic cue to /p t k/, albeit with smaller values than English (cf. Alves et al. 2008, Ahn 2018). Based on these patterns, it is expected that domain-initial strengthening would be observed in the current sample through VOT by showing the largest differences between prosodic contexts in English, followed by Portuguese, and the smallest differences in Spanish (see Figure 1).

RQ3: How does domain-initial strengthening affect unstressed syllables in prenuclear position within the Intonational Phrase?

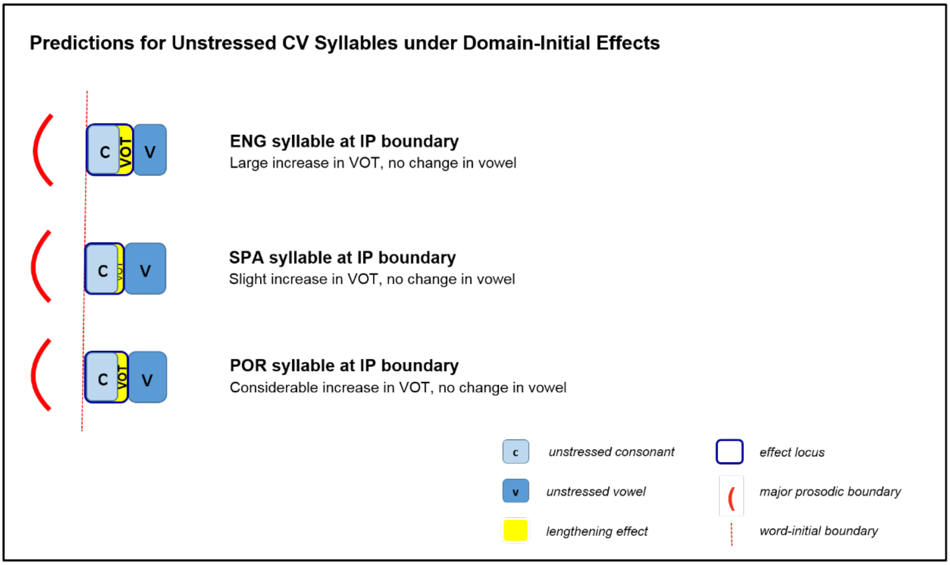


Figure 1: Schematic representation of the expected effects of domain-initial strengthening on unstressed syllables at an IP boundary in the languages in the sample, based on previous findings.

As reviewed above, a great number of studies investigating domain-initial strengthening failed to control for lexical or phrasal prominence (but see Kim et al. 2018). As such, it remains to be determined how the marking of initial prosodic boundaries influences the segmental makeup of unstressed syllables in prenuclear position. One general hypothesis guiding the current investigation is that any effects of boundary marking would be more apparent in those syllables.

If the findings in previous research hold, one would observe boundary-induced changes only to the segment following the IP edge, for instance the consonant in a CV syllable (see Figure 1). The vowel in such syllables would thus manifest the same language-specific characteristics of unstressed vowels near the IP edge as in the middle of it, for instance shorter duration and/or centralization. As mentioned above, the interplay of stress-related reduction on the vowel and boundary-induced strengthening on the consonant in languages like English or Portuguese could potentially lead to changes in the syllable itself, for instance through emerging devoiced vowels due to increased gestural overlap (cf. Jun & Beckman 1994, Mo 2007 for Korean, Davidson 2006 for English, Delforge 2008 for Spanish).

In sum, the current study expands on previous findings by addressing the following key points:

1. Separating domain-initial effects from lexical prominence by investigating word-initial unstressed syllables.
2. Controlling the placement of nuclear accent in the target sentences through elicitation of narrow focus away from the initial boundary.
3. Avoiding an excessive influence of individual idiosyncrasies on the overall results by recording a larger group of speakers per language.
4. Evaluating the role of language-specific phonology by performing a direct comparison of three languages using similar materials.

2 Experimental materials and methods

This study investigated domain-initial effects on fully unstressed syllables in prenuclear words in three languages with contrastive lexical stress. The method presented here focuses on speech data sampled under experimental control. Separate reading tasks were conducted with native speakers of American English, Mexican Spanish, and Brazilian Portuguese using the same experimental procedure and analysis methods.

2.1 Participants and experiment procedure

In total, 42 participants, aged between 18 and 31, took part in the reading experiment. Fourteen speakers of each of the three languages make up the sample (English: eight female; Spanish: twelve female; Portuguese: eight female). Participants were university students, mostly undergraduate, and were naïve to the purposes of the study. All participants were native L1 speakers of their respective sample language, with no known vision, hearing, or speech impediments.

All English speakers reported being monolingual; Portuguese speakers were bilingual in English to different degrees. All Spanish speakers were fully bilingual in English, having nonetheless first acquired Spanish in the home from both of their (Mexico-born) parents. Most speakers of each language came from the same state within their respective countries (American English: New Mexico; Mexican Spanish: Chihuahua; Brazilian Portuguese: Minas Gerais). However, no specific efforts were made to control for dialect representation within country varieties, so that any dialectal differences that may have arisen are not accounted for here.

Experiments took place in soundproof or quiet rooms at the University of New Mexico in the United States, and at the Universidade Federal de Minas Gerais in Brazil. The full reading session lasted between 25 and 50 minutes. Acoustic data were acquired at a sampling rate of 16 kHz through an Audio-Technica USB microphone plugged directly into a laptop. The sound editing software Audacity (Audacity Team 2014) was used to record all participants.

Stimuli were presented on sheets of paper with six to eight sentences per page in large fonts, one sentence per line. Participants read each carrier sentence three times in pseudo-randomized order, which were interspersed among filler items eliciting various other types of prosody (e.g. lists, questions, etc.). Additionally, different presentation orders were devised for each language, so that the items were read in the same order by only three or four speakers per language. Reading sessions were divided into two (for English) or three blocks (for Spanish and Portuguese) to avoid experimental fatigue. Speakers were encouraged to read the sentences at their own pace.

English speakers read 120 sentences in total whereas Spanish and Portuguese speakers read 180 each. These included sentences with test words beginning with /m n/ which were intended for future analysis. A short practice run was conducted before the recording started. Unless prompted by participants themselves, the experimenter provided no feedback/corrections during recording sessions. However, whenever there was an error or disfluency, speakers could repeat a given sentence if they wished. The experimenter interacted with participants in their own respective languages.

2.2 Experiment materials

2.2.1 Target syllables and test words

Target syllables consisted of a CV sequence of /p t k/ plus a monophthong in unstressed, word-initial position. For English, the vowel in the nucleus of the target syllable was always /ə/, selected to avoid possible confounds of secondary stress (cf. Crystal & House 1988, de Jong 2004). In both Spanish and Portuguese, target syllables had high and low vowels (i.e. /i a u/). The use of two different high vowels in the Romance languages was due to language-specific constraints in the distribution of stop plus high vowel in those languages. Vowel height was then included in models as dependent variables due to differences in duration in high and low vowels in both Spanish and Portuguese (cf. Hualde & Nadeu 2014, Cristófaro-Silva et al. 2019). Test items were trisyllabic words with penultimate stress in all three languages. Using three-syllable words reduces the possibility of secondary stress due to rhythmic constraints. Most test words had the overall /CVˈCV.CV/ shape. Immediately following target syllables in all test words was the stressed syllable, which always started with a voiceless stop or a voiceless affricate. A total of 6 English words, and 12 Spanish and 12 Portuguese words were tested in this investigation. Table 1 illustrates the test words used in the experiment by the consonant in the word-initial syllable (see Appendices A–C for the full list).

Table 1: Examples of target words used in the English, Spanish, and Portuguese stimuli. Note that the sequence [ti] is absent in most dialects of Brazilian Portuguese

	Segment		
Language	/p/	/t/	/k/
English	pʰəˈtʰɪʃənz ‘petitions’	tʰəˈkʰilə ‘tequila’	kʰəˈpʰeɪʃəz ‘capacious’
Spanish	piˈkete ‘injection prick’	tiˈpexo ‘an idiot’	kiˈteɲo ‘person from Quito’
Portugese	pɪˈtadə ‘a pinch of’	tʊˈtɛlə ‘guardianship’	kɪˈtadə ‘paid-off’

2.2.2 Experimental conditions and carrier sentences

This study uses the IP as test ground for its hypotheses. Test words were tested in two different conditions in carrier sentences: either at the beginning of an IP (IP-initial condition) or in the middle of the IP (IP-medial condition).[2] Thus, word-initial target syllables were themselves either IP-initial or IP-medial. Test words always occurred in prenuclear position. Overall carrier sentence size was kept at 25 canonical syllables in each of the three languages, for both conditions. This precaution was taken as a control for utterance length effects on articulation rate, and thus segment duration (cf. Fónagy & Magdics 1960, among others).

Following Cho & McQueen (2005), the IP containing the target syllable was always preceded by a precursor IP to ensure that the target syllable necessarily occurred at the beginning of the IP domain rather than at the Utterance. In turn, the position of the target syllable within the test IP was kept constant for each

---

[2] One could alternatively call the two conditions “IP-initial” and “Wd-initial”. However, given that target syllables were always word-initial, the “Wd-initial” condition is referred to as “IP-medial” throughout the chapter to avoid confusion.

condition across languages. In the Spanish and Portuguese stimuli, the target syllable always occupied the seventh slot in the IP-initial condition, whereas in the English sentences it was always the sixteenth in the carrier utterance. Punctuation marks (i.e. colons, semicolons, or commas) elicited a separation between the precursor and test IPs (cf. Turk & Sawusch 1997, Keating 2004).

A further series of measures were undertaken to guarantee that test words did not receive nuclear pitch accent. This is because in all three languages, nuclear pitch accent has been found to increase duration in words/syllables that bear the main IP accent (e.g. Cantoni 2013 for Brazilian Portuguese; Hualde & Prieto 2015 for Spanish). Contrastive or corrective narrow focus was thus elicited elsewhere in the IP in the IP-initial condition. In the three languages, focused words receive the nuclear pitch accent.

In each carrier sentence in the IP-initial condition, the precursor IP(s) established the context for narrow focus in the IP that contained the test words. The exact number of precursor IPs or words in them varied due to differences in word size and syntactic factors between the languages. In addition to contextual information, focused words were capitalized and participants were instructed to place emphasis on them (see Turk et al. 2006 for a discussion of this strategy). Table 2 illustrates the carrier sentences used in the study (see Appendices A–C for the full set of stimuli, with translations).

Additionally, carrier sentences consisted of a variety of meaningful passages, so as to avoid inadvertently eliciting contrastive focus on test items which occurs in sentences of the "Say x again" type. Finally, in order to avoid monotonous intonation due to experimental fatigue (Xu 2010), filler items consisted of a variety of sentence types of different sizes, including yes-no and wh-questions, lists, exclamations, and contrastive focus sentences (see Appendices A–C). All filler sentences were obtained from the Brigham Young University (BYU) corpora of the specific varieties of the languages investigated: Corpus do português (Davies & Ferreira 2006), Corpus del español (Davies 2016), and the Corpus of Contemporary American English (Davies 2008).

2.3 Variables

The acoustic marking of domain-initial position was evaluated separately for each segment, so that there are two broad types of dependent variables: consonant measures, and vowel measures. Independent control variables include phonological as well as sociolinguistic variables (i.e. age, gender). These are described in turn below.

Table 2: Sample carrier sentences for test words ‘capacious’, *pitada* ‘whistle blow’/‘pinch (e.g. of pepper)’ in Spanish and Portuguese, respectively. The sentences below show how unstressed CV syllables (e.g. kə in ‘capacious’ [kʰəˈpʰeɪʃəz]) were tested under the two experimental conditions in this study. The IP-initial boundary is represented by <||>; the IP-medial boundary is represented by <#>

Cond.	Lang.	Carrier sentence
IP-initial		
	ENG	It doesn’t refer to ability! You can check for yourself || capacious means ROOMY or full of space
	SPA	‘Estás confundido || pitada quiere decir SOPLADO más que sonido o pitido
	POR	Tem muito sal aqui, || pitada quer dizer só um POUQUINHO do ingrediente na receita
IP-medial		
	ENG	It is very sad there’s not too much they can do at this point: the city’s # capacious museum closed
	SPA	A causa de la lluvia, el árbitro Federico dió la # pitada a las tres horas
	POR	Pimenta caiena é mais forte do que do reino, só uma # pitada tá mais que bom

2.3.1 Variables pertaining to /p t k/ in target syllables

2.3.1.1 Voice onset time (VOT)

Defined here as the interval immediately after the release of the stop up to the onset of voicing (Lisker & Abramson 1964). VOT was measured manually from first peak of the stop burst release up to the zero crossing nearest the onset of the second formant in the following vowel, as shown in the spectrogram. In case of multiple release bursts, the first burst was used. In the absence of clear burst releases, periods of visible aspiration in the spectrogram were also measured as VOT (cf. Abramson & Whalen 2017), in which case the beginning of the aspiration noise was taken as the acoustic delimiter for segmentation.

2.3.1.2 Occurrence of stop release burst(s)

Defined here as a transient noise pulse at the release of the built-up air pressure during the voiceless stop closure. The occurrence of a stop release burst was measured by visual inspection of both waveform and spectrogram.

2.3.2 Variables pertaining to /ə/ or /i a u/ in target syllables

2.3.2.1 Vowel duration

Vowel duration was measured from the onset of the vowel's F2 to its offset as seen in the spectrogram and the waveforms.

2.3.2.2 First formant (F1)

F1 was extracted from the midpoint of the target vowel as labeled in the TextGrid. F1 data were subsequently normalized for each token based on the means and standard deviations calculated over all productions by the same speaker.

2.3.2.3 Fundamental frequency (F0)

Fundamental frequency may serve as a correlate of lexical prominence. For all tokens produced by female speakers, the range of analysis for F0 was specified as 100–400 Hz. Tokens from male speakers were analyzed using a range of 50–250 Hz. If domain-initial strengthening shows no influence on F0, values would not differ significantly between conditions.

2.3.2.4 Vowel dispersion

Vowel dispersion was measured only in the Spanish and Portuguese data. It is defined as the location of vowels along the back-front and high-low dimensions measured as a function of F1 and F2. As with F1, F2 was extracted from the midpoint of the vowel as labeled in the TextGrid.

2.3.3 Independent control variables

2.3.3.1 Silent interval

Silent interval is defined here as the interval without vocal fold vibration in the waveform, in milliseconds. For sentences in the IP-initial condition, silent interval duration was measured from the end of the last segment of the last word in the precursor IP up to the beginning of the first segment of the test word (as generated by the automatic forced aligner). Silent interval duration was measured in the IP-medial condition using the same criteria as in the IP-initial condition. It should be noted that the automatic forced aligners might include part of the closure of the voiceless stop in its measure of the putative pause that occurs before the test word.

2.3.3.2 Duration of the stressed syllable in the test word

The stressed syllable is defined as the most prominent syllable in a word, as specified in the lexicon of each of the three languages. In the Spanish and Portuguese data, stressed syllables were measured based on the output of the purpose-designed Praat script. In the English data, the stressed syllable was measured manually from the F2 offset of the unstressed vowel in the target syllable up to the offset of F2 in the vowel in stressed syllable for all but one test word. In the case of the word ‘patrolmen’, the only test word whose stressed syllable ended in a coda consonant, the syllable was measured from the F2 offset of the unstressed vowel /ə/ up to the beginning of the nasal murmur of the /m/ in the final syllable /mɪn/. Stressed syllable duration values were log-transformed for analysis given differences in the number and types of segments in the stressed syllable across the different words used as stimuli (see Section 2.2).

2.3.3.3 Articulation rate

Defined as the number of syllables divided by phonation time, excluding silent intervals over 200 milliseconds. Articulation rate was measured automatically through a Praat script (de Jong & Wempe 2009), which estimated the number of syllables based on acoustic information in the audio files containing individual carrier sentences. The script identifies syllable nuclei by detecting peaks in intensity (dB) that occur between two dips in the audio file, thus avoiding measuring segments other than vowels. It is assumed that the faster the articulation rate, the shorter the acoustic durations (cf. Fónagy & Magdics 1960, Crystal & House 1988, see also Kessinger & Blumstein 1998 specifically for VOT).

2.3.3.4 Place of articulation

Place of articulation has been shown to influence VOT values (Cho & Ladefoged 1999, for crosslinguistic data; Avelino 2018, for Mexican Spanish; and Ahn 2018, for Brazilian Portuguese), as well as stop burst release (Winitz et al. 1972).

2.3.3.5 Vowel height

Vowel height constituted a variable only in the analyses of Spanish and Portuguese. Vowel height was coded as high, or low, based on citation forms of the vowels in target syllables. High vowels were /i u/, and the low vowel consisted of /a/ alone.

2.3.3.6 Repetition

Each speaker read the stimuli three times. This variable identifies the order of the individual productions of each target word in the reading task: first, second, or third, for each speaker. Previous results (e.g. Fowler & Housum 1987) suggest that segment duration in the second or third productions will be shorter than the first occurrence in the stimuli.

2.3.3.7 Word frequency

Word frequency was operationalized as the number of occurrences of target words per million in the Corpus of Contemporary American English (Davies 2008), the Corpus del español (Davies 2016) for Spanish, and the Corpus do português (Davies & Ferreira 2006), for Portuguese. These values were log-transformed for the statistical analysis. Frequency of occurrence of a lexical item correlates with its overall duration (Bell et al. 2009), so that the higher the frequency, the shorter the duration.

2.4 Specific criteria for confirming IP-initial boundaries

This study compares words produced at IP-initial boundary against those produced phrase-medially. Hence, it was crucial that the production of carrier sentences matched the prosodic context they were designed to elicit. The presence of a long silent interval (i.e. 200 ms or more) between the precursor and the test IP was used as the primary criterion for determining the occurrence of a prosodic boundary. Silent intervals are particularly relevant in the current study since they can serve as indications of prosodic boundary strength (see Krivokapić 2014 for a review). Specifically, there is evidence to suggest that the duration of a pre-boundary pause may correlate with gestural magnitude of the first segment following the pause (e.g. Beňuš & Šimko 2014, see also Ramanarayanan et al. 2009).

Whenever the silent interval between precursor and test IP was shorter than 200 milliseconds, two additional acoustic parameters were used as a confirmation that test words in the IP-initial condition were in fact produced at the left edge of the phrase: pitch declination and reset, and/or the presence of creaky voice (i.e. "phrase-final creak", see Garellek 2015) in the last word in the precursor IP. Both parameters were assessed by visual inspection of the waveforms and spectrograms.

Pitch declination and reset is defined as a lowering of the pitch range between the early part of the precursor IP and the end of that IP, without regard to the tonal description (e.g. HL% or L%). Such lowerings were always followed by a reset of the pitch excursion, meaning that the speaker reset their pitch at the start of the test IP at a higher level than that of the precursor IP. Pitch declination and reset have been described as a cue to IP boundaries in broad statements in all three languages in the sample (Pierrehumbert & Hirschberg 1990, for English; Frota et al. 2007, for Spanish and Portuguese).

In turn, phrase-final creak is defined as a stretch of the speech signal characterized by irregular (e.g. less periodic) F0 and amplitude changes (Redi & Shattuck-Hufnagel 2001, Garellek 2015) that occurs at or near the end of a prosodic phrase. Phrase-final creak has been found to correlate with the end of larger prosodic domains in English (Redi & Shattuck-Hufnagel 2001), Spanish (de la Mota et al. 2010), and Portuguese (Frota & Moraes 2016) alike. Creaky voice was noted as phrase-final creak when it affected all voiced segments in the last word of the background sentence (Garellek 2015) for English and at least the last syllable for Spanish and Portuguese.

2.5 Data extraction, data exclusion

Acoustic measurements were done in Praat (Boersma & Weenink 2019). The FAVE-align automatic forced aligner (Rosenfelder et al. 2014) was used to generate segment intervals in TextGrids for the analysis of the English stimuli, which were then hand-corrected as needed. Syllable, word, and phrase intervals were created manually based on the output for segments generated by FAVE. For the Spanish and Portuguese datasets, the automatic forced aligner EasyAlign (Goldman 2011) generated syllable, word, and phone intervals. Hand-corrections were made where needed. Subphonemic segmentation was done manually based on the acoustic information available in the spectrogram and waveform. Prosodic annotations were also done manually. All annotated data were extracted automatically from the TextGrids using purpose-designed Praat scripts.

Each speaker produced 36 tokens of the test words (6 words × 2 conditions × 3 repetitions), totaling 504 tokens per language prior to inspection. Tokens produced with unexpected prosodic or intonational patterns, such as laughter, hesitancy, or misplaced nuclear accent on test words, were excluded from analysis. Errors affecting the precursor IP also led to exclusion, although sentences containing errors affecting words that followed the test word were kept. Finally, test words in English or Portuguese showing vowel deletion in the target syllable were also excluded from the acoustic analyses (50 English tokens, and 31 Portuguese tokens). In total, 423 tokens of English data were analyzed. For Spanish, the data correspond to 413 tokens, whereas Portuguese results derive from 400 tokens.

2.6 Statistical analyses

All statistical analyses were conducted in R (R Core Team 2018). Visual inspection of the data involved generating basic graphs that displayed broad distribution patterns of the dependent variables across experimental conditions. The Shapiro-Wilk normality test was then applied to variables using the generic function built in R as a way to assess whether values in numeric variables followed a normal distribution. For instance, the Shapiro-Wilk test confirmed that VOT, unstressed vowel duration, and F1 data failed to reach a normal distribution in all three languages. The two-tailed Wilcoxon non-parametric statistical hypothesis test was then applied separately to the three continuous dependent variables in each language to determine whether prosodic context yielded statistically significant differences in the data (see Wilcoxon tests results). Variables that showed no statistical difference between conditions were excluded from further analyses of the individual languages (see Results). Mixed effects models were only fit for variables that showed statistically significant differences between experimental conditions.

Numeric variables such as duration and VOT were log-transformed. Additionally, numeric predictors were z-scaled using the generic function `scale()` in R (cf. Gries 2013, Bell et al. 2009). At this point, any remaining outliers (i.e. figures that were three median absolute deviations away from the median) were further excluded from the subset. Linear mixed-effects models were then fit to each subset of the data using the `mixed()` function in the Afex package (Singmann et al. 2015) with all appropriate independent variables as predictors. The `mixed()` function also produces *p*-values for the likelihood ratio test.

Variables in each model were introduced through a backward selection procedure to help guard against model overfitting. Following this procedure, the first model was fit with all individual predictors and theoretically relevant interactions. Interactions tested included: articulation rate vs. repetition, duration of the stressed syllable vs. condition, place of articulation vs. VOT, vowel height vs. vowel duration (for Spanish and Portuguese only), and vowel duration vs. repetition. The interaction of duration of the stressed syllable and condition specifically tested whether proximity to the IP-initial boundary influenced stressed syllable duration. After each model was fit, it was compared to a set of models with one fewer predictor via the generic function `anova()` in R. The Akaike Information Criterion (AIC) was used as a goodness-of-fit measure for model comparison. The predictor that contributed the least to model fit was then removed from the full model. The process was repeated until the final model was significantly better than all possible alternatives with one fewer predictor.

3 Results

This study investigated how the marking of domain-initial boundaries affects the acoustic properties of segments in fully unstressed syllables in three lexical stress languages. The following results stem from comparing test words measured under two experimental conditions: IP-initial, and IP-medial. Results are presented first as an overview, followed by detailed descriptions of the individual languages (see also Statistical summaries and Wilcoxon tests results for statistical data).

3.1 Crosslinguistic summary

Table 3 summarizes the results of the comparison between prosodic contexts in the three languages in the sample. All languages showed differences between IP-initial and IP-medial contexts, although the specific acoustic correlates varied somewhat between the languages.

Table 3: Comparison of results by prosodic condition in the three languages (significance levels are *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; n.s.: not significant; IP: IP-initial; Wd: Word or IP-medial).

Measure/Language	English	Spanish	Portuguese
VOT /p t k/	IP > Wd *	n.s.	n.s.
Vowel duration	n.s.	IP > Wd ***	IP > Wd ***
F1	n.s.	n.s.	n.s.
F0	n.s.	IP > Wd ***	n.s.
Vowel dispersion	NA	n.s.	n.s.
Burst release	IP > Wd **	IP > Wd ***	IP > Wd **
Stressed syllable duration	IP < Wd ***	IP < Wd ***	IP > Wd ***

As shown in Table 3, the languages in the present study show some similarities and quite a few differences in how boundary-initial marking affects the acoustic properties of trisyllabic words with penultimate stress. All languages showed more burst releases at stop closure for /p t k/ that followed the IP-initial boundary than IP-medially. On the other hand, VOT is used to mark stops in unstressed syllables only in English, with Spanish and Portuguese /p t k/ failing to show significance in how VOT lag differs between prosodic contexts (see Figure 2). Additionally, whereas tokens of English /ə/ in the target CV syllable were unaffected by boundary marking, the duration of Spanish and Portuguese /i a u/ under similar conditions shows significant effects of prosodic context (Figure 3). Spanish words, in particular, also showed higher F0 values near the IP boundary.

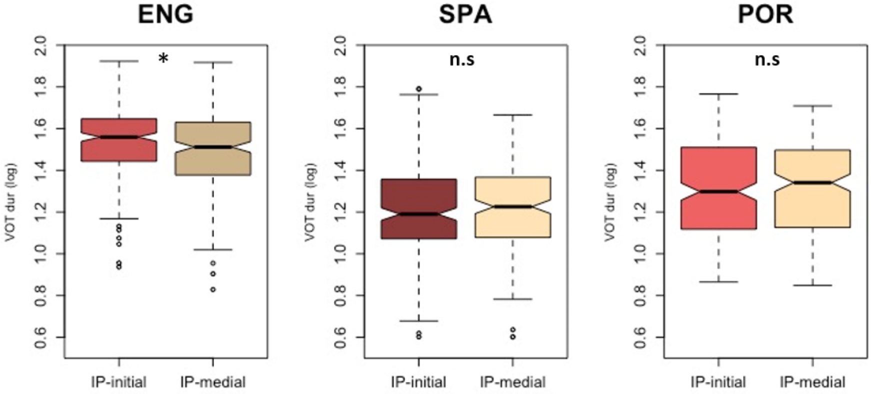


Figure 2: Duration of VOT lags in target syllables (log values). Box plots show VOT duration results measured in target syllables across the stimuli (e.g. English "capacious", Spanish/Portuguese *pitada*), as a function of prosodic context. IP-initial values in left box, IP-medial data in right box; y-axis shows the same scale for the three languages; *p<.05; n.s.: not significant.

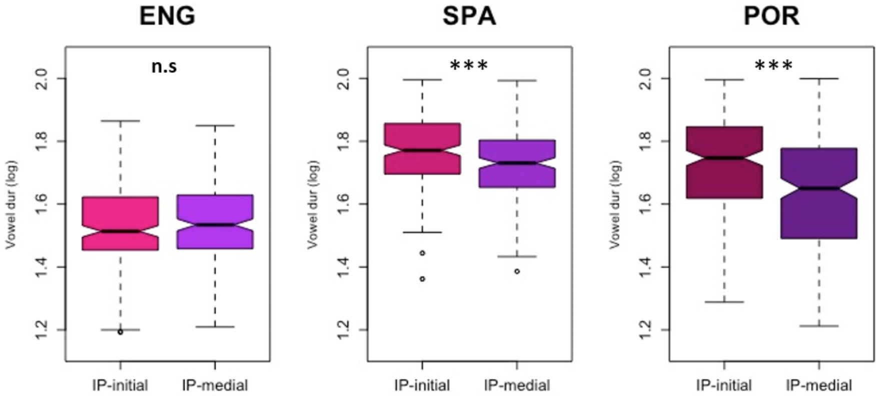


Figure 3: Duration of vowels in target syllables (log values). Box plots show vowel duration results measured in target syllables across the stimuli (e.g. English "capacious", Spanish/Portuguese *pitada*), as a function of prosodic context. IP-initial values in left box, IP-medial data in right box; y-axis shows the same scale for the three languages; ***p<.001; n.s.: not significant.

What's more, while the non-boundary adjacent stressed syllable shows systematic variation in all three languages, such an effect is uneven; English and Spanish words pattern in one way, and the Portuguese data in another. In both English and Spanish, the stressed syllable was shorter near the IP boundary, whereas in Portuguese it was longer. It should be noted, however, that stressed syllables were control variables, so that their phonetic content in the current data was not controlled for. The following sections detail these findings in light of the statistical techniques used to evaluate them.

3.2 Consonant results: unstressed /p t k/ in target syllables

3.2.1 Burst release at stop closure

The occurrence of a burst release at the stop closure was evaluated for each language via Fisher's exact tests. Based on Lavoie (2001), it was expected that consonants at an IP-initial boundary would show more burst releases than those occurring IP-medially. This prediction was borne out in the data, although it is noteworthy that there was an overall high rate of burst releases for all consonants in the sample. Table 4 shows the burst release results for the three languages.

Table 4: Proportions of occurrence of burst at /p t k/ release in the sample by prosodic context; *N* = number of tokens. Significance levels are *** $p < 0.001$; ** $p < 0.01$

	English**		Spanish***		Portuguese**	
	Yes (*N*)	No (*N*)	Yes (*N*)	No (*N*)	Yes (*N*)	No (*N*)
IP-initial	1.0 (222)	0	1.0 (221)	0	1.0	0
IP-medial	0.95 (200)	0.05 (10)	0.94 (11)	0.03 (11)	0.97 (176)	0.03 (6)

3.2.2 VOT in English

VOT values were evaluated through mixed effects models with word, and speaker as random effects. As mentioned above, English was the only language in the current study that showed statistically significant differences for VOT values in target syllables between prosodic conditions. A total of 432 observations related to VOT in voiceless stops were entered in the English mixed-effects model. The model was fit through backward selection of variables, meaning that predictors that failed to improve model fit (i.e. $p \geq 0.05$) were excluded from the analysis.

The best model for VOT (log-transformed) as the response variable included speaker and word as random effects, and the following variables as fixed effects: prosodic context (two levels: IP-initial or IP-medial), consonant type (three levels: /p/, /t/, or /k/), and the interaction of duration of the stressed syllable and prosodic context. The R function `mixed()` automatically calculated p-values. The final fixed and random effects estimates appear in Tables 5 and 6, respectively.

Table 5: Main effects for the English model with VOT in unstressed word-initial /p t k/ as the response variable (log-transformed). The reference levels for categorical predictors are IP-initial for prosodic context and '/k/' for consonant type. Significance levels are $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p > 0.01$; n.s.: not significant.

Estimate	β	SE	t	$p(t)$
Intercept	1.900	0.397	4.78	0.004**
Prosodic context = IP-medial	−1.831	0.588	−3.11	0.002**
Stressed syllable duration x IP-medial	0.642	0.214	3.00	0.009**
Consonant = /p/	−0.247	0.028	−8.80	0.166 n.s.
Consonant = /t/	−0.059	0.028	−2.11	0.389 n.s.
Stressed syllable duration x IP-initial	−0.127	0.172	−0.74	0.492 n.s.

Table 6: Random effects intercept – English VOT in target syllables (SD: standard deviation)

Variable	Variance	SD
Speaker	0.0088	0.0941
Word	0.0002	0.0126
Residual	0.0445	0.2110

As summarized in Table 5, the results from the generalized linear model revealed significant main effects of prosodic context, with the duration of the VOT lag being shorter at an IP-medial boundary than at an IP-initial one. The model also showed an effect of the interaction of duration of the stressed syllable and prosodic context: the longer the stressed syllable, the longer the VOT at an IP-medial boundary.

3.3 Vowel results

As mentioned above, variables pertaining to unstressed vowels only showed significant differences between contexts in Spanish and Portuguese (see Statistical summaries for English vowel results). Of these, duration was significantly different between IP-initial and IP-medial conditions in both Spanish and Portuguese, whereas F0 was significant for Spanish alone. F1 values failed to reach significance in any language (see Statistical summaries for all results).

3.3.1 Vowel-related results in Spanish

Vowel duration values were evaluated through mixed effects models, with speaker and word as random effects. A total of 413 observations related to unstressed vowels in the Spanish data were analyzed. The best model for Spanish vowel durations (log-transformed) as the response variable included the following variables as fixed effects: duration of silent interval, word frequency, vowel height (two levels: high or low), and the interaction of duration of the stressed syllable and prosodic context (two levels: IP-initial or IP-medial). The `mixed()` function in R automatically calculated *p*-values. The final fixed and random effects estimates appear in Tables 7 and 8, respectively.

Table 7: Main effects for the Spanish model with vowel duration in target syllable as the response variable (log-transformed). The reference level for vowel height is "high". Significance levels are $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; n.s.: not significant.

Estimate	β	SE	t	$p(t)$
Intercept	1.388	0.018	10.89	$<0.001^{***}$
Duration of pause	0.020	0.008	4.71	$<0.001^{***}$
Vowel height = low	0.099	0.010	8.38	$<0.001^{***}$
Stressed syllable duration x IP-initial	0.172	0.072	2.39	0.017^{*}
Word frequency	−0.013	0.006	−2.22	0.01^{*}
Stressed syllable duration x IP-medial	−0.004	0.071	2.40	0.659 n.s.

As summarized in Table 7, results from the generalized linear model revealed significant main effects of vowel height, with the duration of the low vowel /a/ being overall longer than /i u/. The model also showed that a longer silent interval increases the duration of the vowel in the post-boundary CV syllable. Because silent pauses occurred most consistently before test words in the IP-initial

Table 8: Random effects intercept for Spanish /i a u/ duration (SD: standard deviation)

Variable	Variance	SD
Speaker	0.0088	0.0560
Word	0.0002	0.0050
Residual	0.0445	0.0853

condition, one can interpret the main effect of silent interval as an indirect correlation of domain boundary level with duration of unstressed vowels in the Spanish target syllables. There was also an effect of the interaction of duration of the stressed syllable and prosodic context: the longer the stressed syllable, the longer the vowel in the unstressed syllable at a prosodic boundary. Finally, the word frequency of the test word also showed the expected influence on duration: the higher the frequency, the shorter the unstressed vowel.

3.3.2 Vowel duration in Portuguese

In total, 400 tokens of Portuguese vowel duration data were fed into the mixed effects model. The model was fit through backward selection using all the applicable variables. Once again, predictors that failed to improve model fit (i.e. $p \geq 0.05$) were excluded one at a time until the model described here was finalized. The best model for vowel duration (log-transformed) as the response variable in the Portuguese data included speaker and test item as random effects, and the following variables as fixed effects: duration of silent interval, vowel height (two levels: high or low), and the interaction of duration of the stressed syllable and prosodic context (two levels: IP-initial or IP-medial). The `mixed()` R function automatically calculated p-values. The final fixed and random effects estimates appear in Tables 9 and 10, respectively.

Results from the generalized linear model shown in Table 9 revealed significant main effects of vowel height, with the duration of the low vowel /a/ being overall longer than /i u/, similarly to the findings for Spanish. The model also showed that a longer pause duration is associated with increased duration of the vowel in the post-boundary CV syllable. There was also an effect of the interaction of duration of the stressed syllable and prosodic context: the longer the stressed syllable, the longer the vowel in the unstressed syllable at a prosodic boundary, more at an IP boundary than at the IP-medial domain.

Table 9: Main effects for the Portuguese model with vowel duration in the unstressed syllable as the response variable (log-transformed). The reference level for vowel height is "high"

Estimate	β	SE	t	$p(t)$
Intercept	2.464	0.419	17.88	<0.001***
Vowel height = low	1.591	0.109	10.70	<0.001***
Duration of pause	0.037	0.005	7.70	<0.001***
Stressed syllable duration x IP-initial	0.025	0.008	3.29	0.003**
Stressed syllable duration x IP-medial	0.018	0.010	1.83	0.298n.s.

Table 10: Random effects intercept – Duration of /i a u/ in target syllables (SD: standard deviation)

Variable	Variance	SD
Speaker	0.0018	0.0424
Word	0.0003	0.0164
Residual	0.0115	0.1071

4 Discussion

4.1 Implications for our understanding of prosodic structure

The current study is novel in two main ways: first, it provides evidence for the acoustic expression of domain-initial boundaries using a large sample. Secondly, by using the same materials to assess the phenomenon in three languages, it provides a direct evaluation of the claim that domain-initial boundaries manifest themselves in language-specific ways (e.g. Cho & McQueen 2005). Overall, this study corroborates the general hypothesis that words occurring immediately after a major prosodic boundary differ in their phonetic properties from those that follow a lower level boundary (Fougeron & Keating 1997). Furthermore, the different analyses presented above show that domain-initial strengthening operates on the phonetic properties of word-initial unstressed syllables, expanding them in language-specific ways.

As such, these results partially confirm two of the hypotheses that guided the current study (see Section 1.3). First, that domain-initial strengthening is not limited to articulation, and that acoustic variables are useful tools to describe (IP-) initial boundaries. Secondly, that different languages show particular correlates of domain-initial strengthening. Within target CV syllables, English showed effects on the boundary-adjacent segment alone, whereas Spanish and Portuguese showed acoustic differences on both the consonant (i.e. occurrence of burst at stop release) and on the vowel (e.g. duration).

On the other hand, these language-specific characteristics manifested themselves differently from what had been predicted. Instead of differences in the magnitude of the effect, the current findings showed that both the type of segments affected (i.e. consonants in English, vowels in Spanish and Portuguese), and how those segments were affected (e.g. F0 in Spanish vowels, dispersion in Portuguese ones) differed. As such, these results reveal a somewhat inconsistent behavior of the variables under study. While these findings can be taken as confirmation of the phonological specificity of domain-initial strengthening, they render generalizations made over the entire sample much less straightforward. These and further limitations are taken up in more detail in Section 4.5 below.

The results presented here differ from those in previous literature in two relevant ways. First, domain-initial effects extended beyond the segment immediately following the boundary in Spanish and Portuguese, in which the vowel in the target CV syllable was longer in the IP-initial condition than IP-medially. Secondly, in the three languages, the stressed syllable, which was not boundary-adjacent, showed significant durational differences between prosodic contexts. Taken together, these results not only contradict the hypotheses laid out in Section 1.3, but also go against previous results suggesting that domain-initial strengthening effects are limited to the very first segment following the major boundary (Byrd et al. 2006, Cho & Keating 2009, Bombien et al. 2010, among others). As mentioned above, one key aspect of the current investigation is that it controlled for various levels of lexical stress and prosodic prominence. It is thus possible that the differences found here relate to the issue of prominence, taken up in more detail below.

4.2 Lexical stress and the locality of domain-initial strengthening

This study was designed to isolate the influence of lexical stress and phrasal prominence from domain-initial strengthening. While this investigation focused on the segments in the boundary-adjacent unstressed syllable, the following stressed syllable in test words was measured as a control variable. The diagram in Figure 4 depicts the findings of the study for target unstressed syllables (represented as <cv>) while also showing the stressed syllable (i.e. <’CV>) in test words.

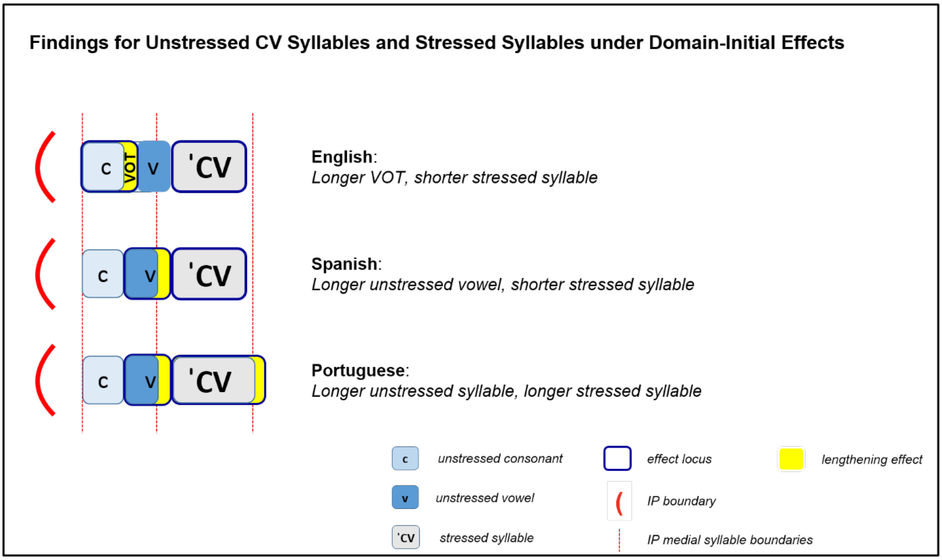


Figure 4: Schematic representation of the effects of boundary marking on IP-initial unstressed syllables, and the following stressed syllable for the languages in the study. Dashed lines represent syllable boundaries in the IP-medial condition as reference.

Individual statistical analyses for the three languages revealed that the stressed syllable had a significant main effect on the duration of VOT for English, and on vowel duration in both Spanish and Portuguese. Simply put, the effects observed on segments at/near the IP boundary were linked to the stressed syllable. While novel (to the best of my knowledge) in the domain-initial strengthening literature, these results have parallels in several studies of final lengthening that controlled for stress placement in *phrase-final* words.

The position of the lexically stressed syllable is a decisive factor in determining the scope of final lengthening in English (Kim et al. 2018, Cho et al. 2017, Byrd & Riggs 2008, Turk & Shattuck-Hufnagel 2007, White 2002, Oller 1973), German (Schubö & Zerbian 2023 [this volume]), Spanish (Rao 2010), and Portuguese (Frota 2000), but also in other stress languages such as Estonian (Krull 1997), Greek (Katsika 2016), Italian (Petrone et al. 2014), and Hebrew (Berkovits 1994). These studies and the current findings converge in that they all underscore the importance of lexical prominence in determining the scope of boundary-related effects. The present results are thus compatible with a view of prosodic boundaries and lexical prominence as closely related entities in the expression of prosodic structure (e.g. Turk & Shattuck-Hufnagel 2007, Katsika 2016).

Discussing domain-final boundaries, Turk & Shattuck-Hufnagel (2007) put forward a hypothesis that prosodic lengthening affects both the boundary-adjacent syllable and the stressed syllable that is not immediately adjacent to the boundary. According to this interpretation, the scope of boundary effects is determined by prosodic structure (i.e. the type of domain level) and by the phonological properties of the word (i.e. where stress is located) simultaneously. Put differently, the authors' hypothesis suggests that boundary marking is phonetically expressed with reference to lexical prominence. While articulatory in nature, Katsika's proposal (Katsika 2016) that prosodic boundaries and lexical prominence are integrated could also explain the data obtained in the current study. According to her, prosodic events related to boundary marking (i.e. domain-edge lengthening, articulatory strengthening, phrasal accents, boundary tones, and pauses) are interdependent, with lexical prosody functioning as the interface between phrasal prosody and constriction gestures (Katsika 2016: 169). Katsika's hypothesis is compelling as it can account not only for the acoustic results in the current study, but also for data in studies of domain-final lengthening and phrasal accent.

One possible interpretation of the aforementioned proposals is that phrasal effects would only affect segments in relation to a lexically prominent unit. This approach would imply that prosodic structure is phonetically cued from lower levels up, for instance from the Syllable to the Intonation Phrase. This bottom-up view of the prosodic hierarchy would suggest that smaller domains provide the framework upon which the whole structure is built. In light of the great deal of variability in prosodic phrasing, it may be useful to consider an approach that is more based on the concrete - and perhaps more stable - prosodic properties of lower-level domains such as the Word. Although formulated to account for other types of prosodic phenomena, Turk & Shattuck-Hufnagel (2007) and Katsika (2016) proposals would explain the current findings of lengthening occurring both in the segments immediately following the IP boundary and in the stressed syllable.

Viewed this way, the results of the present study suggest that the lexically prominent syllable may serve as an anchoring point for boundary marking, perhaps in similar ways to how it encodes phrasal prominence. The idea that domain-initial strengthening and lexical stress are interdependent coheres with the existing body of literature showing an association between lexical stress and phrasal accent in terms of pitch movement. In this interpretation, domain-initial effects would begin in the stressed syllable, and move leftwards to the phrase-initial boundary. That is, the locus of domain-initial effects would be best described as the stressed syllable, and the scope of the effect would potentially include segments between that syllable and the major prosodic boundary. Differences in how lexical stress behaves phonetically would then explain specificities found in the implementation of prosodic structure, as suggested by Cho (2016).

4.3 The linguistic function of boundary effects

The evidence gathered in this study suggests that there is more to domain-initial effects than a purely biomechanical motivation. If domain-initial effects derived only from the start-up of articulation after a prosodic break, for instance, different languages would show consistent similarities in the way the prosodic effect operates on given segments such as /p t k/. The current results, as well as the findings from multiple studies reviewed in the Introduction, indicate that that is likely not the case. In other words, domain-initial effects differ in relevant ways from the marking of phrase edges before a prosodic boundary.

Domain-finally, the slowing down of articulators towards the end of the phrase suggests a physiological motivation behind pre-boundary effects such as phrase-final lengthening, or phrase-final creak. These phonetic effects can be interpreted as a reflection of the speaker's planning for the upcoming prosodic break, when most articulators will be at rest. This biomechanical process could then explain why phrase-final lengthening and/or phrase-final creak are crosslinguistically common (cf. Jun 2005, 2014, see also references in Garellek 2015).

On the other hand, the observation that the locus of phrase-final lengthening may relate to a lexically prominent syllable introduces a linguistic foundation for the effect. It is noteworthy that pitch movements that encode phrase-final edges also tend to associate with a linguistically relevant unit, such as a lexically prominent syllable in languages with lexical stress. As mentioned in the above discussion, the results of this investigation provide indication that prosodic boundary marking in English, Spanish, and Portuguese also relates to the lexical stress systems of these languages. The correlation of pre-boundary marking with word prosody thus suggests a possible parallel between the phonetic encoding of both edges of a prosodic domain.

The fact that the phonetic marking of the initial edge seems to relate to the segmental phonology of a language bespeaks a perhaps clearer linguistic motivation for domain-initial effects. Given the relevance of prosodic boundaries in speech recognition (Carlson 2009), an increase in phonemic contrast between neighboring segments at a phrase edge could possibility facilitate the parsing of speech. Because stressed syllables are prominent, it could be argued that they serve as a natural anchoring point for the marking of phrase edges – initial and final alike.

4.4 Implications for phonological change: prestressed vowels in Portuguese

The data obtained here suggest that unstressed vowels are longer near an IP boundary in Portuguese, and that this boundary-related lengthening may prevent unstressed vowel reduction from taking place. These combined results may be useful to explain the asymmetry between prestressed and poststressed vowels in the language. As mentioned in the Introduction, Brazilian Portuguese shows a complex system of vowel qualities that relates simultaneously to lexical stress and syllable position within the word. Figure 5 illustrates the current variation in the expression of vowels in Brazilian Portuguese.

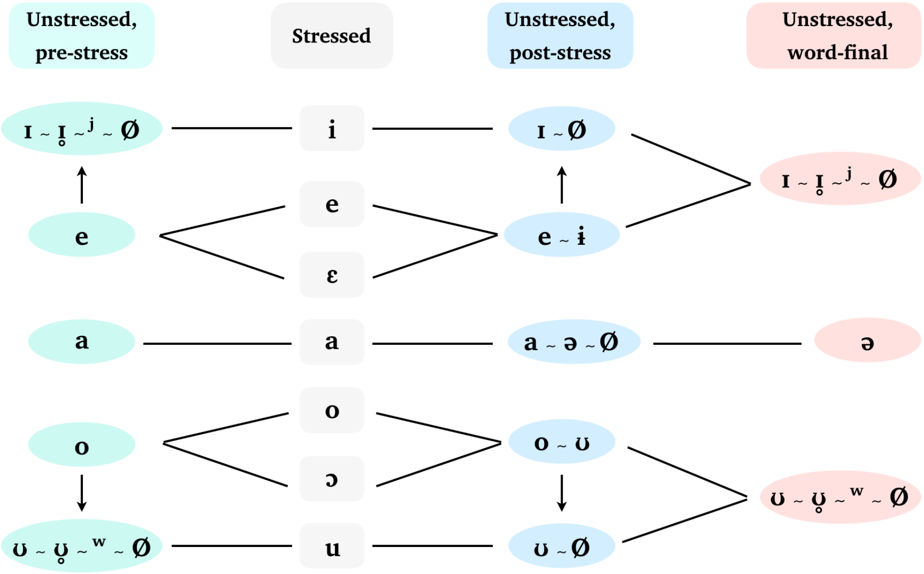


Figure 5: Variation in oral vowels in Brazilian Portuguese given stress placement within the word. Different colors represent variation in unstressed syllables with regards to word boundaries. Common allophonic variation shown within each oval; pointed arrows represent the direction of on-going sound change.

As can be seen in Figure 5, there is a stark contrast between unstressed oral vowels occurring in word-initial as opposed to word-final position. Up to five oral vowels can occur in unstressed word-initial syllables (i.e. /i e a o u/), whereas only /ɪ ə ʊ/ occur in unstressed final syllables. If duration is taken to be one the most important factors in the neutralization of contrasts in vowels, the fact that more vowel qualities are found in prestressed than poststressed position may be a consequence of domain-initial strengthening: the longer duration that can occur in prestressed position facilitates the distinction among more vowel qualities. Although the results regarding vowel dispersion were not significant for the current set of data, there is indication of a trend towards more dispersion at an IP-boundary than IP-medially, depicted in Figure 6.

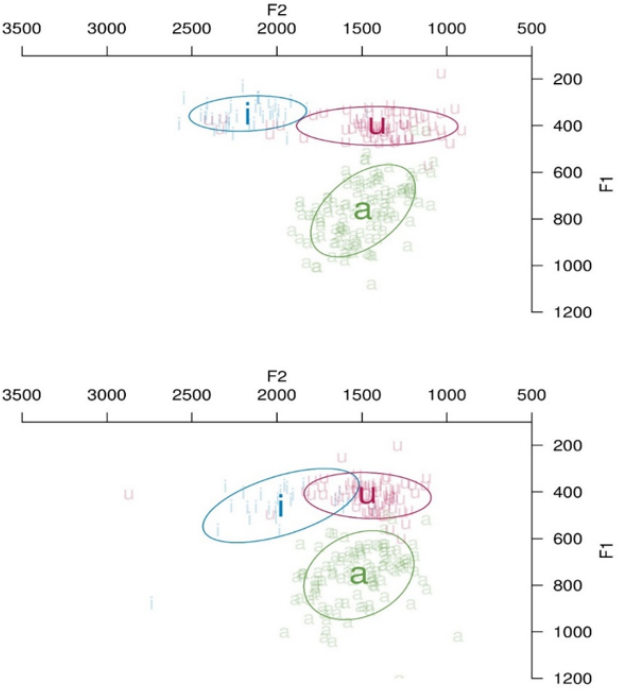


Figure 6: Vowel dispersion in Brazilian Portuguese /i a u/ given adjacency to an IP-boundary in the current study. The top graph shows vowels near the IP-initial boundary; bottom graph shows distribution of vowels IP-medially.

Put differently, the strengthening effect that derives from prosodic structure may be influencing patterns of lexical prominence in Portuguese (cf. Barnes 2006, Scheer & Ségéral 2008 for discussions). As the data (illustrated in Figure 6) suggests, domain-initial strengthening may be partially connected to the still relatively moderate reduction in prestressed syllables compared to poststressed ones in the language. Although more studies are necessary to evaluate this hypothesis, the current findings suggest that prosodic structure is at least a variable that must be taken into consideration in investigations of unstressed vowel reduction in Brazilian Portuguese, and in other languages.

4.5 Limitations and future research

This study focused on the acoustic properties of voiceless stops and vowels immediately following prosodic domain boundaries, compared to sounds that are not adjacent to major boundaries. While some of the results discussed above may appear to contradict earlier articulatory studies, these findings are strictly limited to the acoustic realm, and no claims are made as to the locus, scope or anchor of domain-initial effects in terms of articulation. Future articulatory studies that manipulate lexical stress and accent using similar sample sizes would be useful to reconcile the findings put forward here with previous research on articulation.

Furthermore, more studies are needed to strengthen the present results for consonants other than voiceless stops. Similarly, future investigations of English and Portuguese should also test different vowel qualities in the prestressed CV syllable from the ones used here. For English, the fact that the test syllables always contained /ə/, a vowel which may not always show temporal expansion (see Cambier-Langeveld 1997 for analogous results concerning final lengthening in Dutch /ə/), may have influenced the scope of the boundary effect. For Portuguese, investigating whether proximity to the IP boundary affects the reduction of prestressed /e/ and /o/ to [i ɪ ʲ] and [u ʊ ʷ] would also help determine the extent to which boundary marking has implications to lexical stress reductions.

Since the current project focused on word-initial unstressed syllables, there was much less experimental control for the stressed syllable in the trisyllabic words used in the reading task. A follow-up investigation with tighter experimental control on both the unstressed and stressed syllables could potentially increase the validity of the present findings. As explained above, this study was not explicitly designed to capture differences between stressed syllables, but main effects on stressed syllables were found in all of the languages investigated. More research is hence necessary to confirm the associations found between prosodic context and the phonetic characteristics of stressed syllables not immediately following a major prosodic boundary. A future study could also contrast unstressed and stressed CV syllables like the ones tested here.

Acoustic investigations that manipulate the number of unstressed syllables between the boundary and the stressed syllable would also offer important contributions to our understanding of domain-initial strengthening effects. Additionally, although the test words measured in this research did not receive the main phrasal accent, no specific control was undertaken with regards to the presence and type of prenuclear pitch accents. Future studies that manipulate pitch accent type and placement would constitute a relevant refinement of the methods employed here, including regarding the use of focus statements in both experimental conditions.

Finally, this investigation only considers data from English, Spanish, and Portuguese, all of which are languages with somewhat unpredictable lexical stress. It is possible that a similar acoustic study of languages with other types of word prosody systems, including languages with fixed lexical stress, may result in different associations between boundary marking and lexical prominence. What's more, the claims made here may not be applicable to varieties of these languages other than the ones investigated here, namely American English, Mexican Spanish, and Brazilian Portuguese, given the known prosodic differences between dialects of the same language (e.g. Clopper & Smiljanic 2011: 145, for English; Prieto & Roseano 2010, for Spanish; Frota et al. 2015 for Portuguese). In the case of Spanish, it would be desirable to conduct a similar study using only monolingual speakers instead of the fully bilingual participants recorded for the current project.

5 Conclusion

This study sought to shed new light on the acoustic manifestation of domain-initial strengthening, a type of boundary-induced prominence, from a crosslinguistic perspective. In doing so, it sought to provide more data on the ways through which prosodic structure organizes speech. In contrast to most previous investigations on the topic, the current project looked at unstressed segments occurring in words that did not bear the main (nuclear) phrasal accent. The data obtained from 42 speakers of English, Spanish, and Portuguese revealed that the acoustic correlates of boundary marking extend beyond the initial segment in unstressed CV syllables, affecting the vowel in Spanish and Portuguese, and the stressed syllable in all three languages.

The data discussed here suggest a close connection between the grouping and prominence functions of prosody, in which the stressed syllable may serve as the anchoring point for boundary marking. This proposal is in line with findings from studies of domain-final effects that controlled for stress placement in test words (Turk & Shattuck-Hufnagel 2007, Cho et al. 2013, Katsika 2016). These investigations show that phrase-final lengthening is initiated farther away from the boundary in polysyllabic words that do not have stress on the final syllable. The combined evidence seems to suggest that lexically stressed syllables play a role in marking both domain-initial and domain-final boundaries of major phrases. This function of the stressed syllable would then add to its already established function in marking phrase-level prominence in some languages, and thus provide support to the view that prosodic structure manifests itself phonetically through the interaction of segmental and suprasegmental factors.

More broadly, the results of the current study reinforce the idea that speakers actively use language-specific phonological knowledge (e.g. VOT lag for English /p t k/, vowel duration for Spanish and Portuguese /i a u/) to implement phonetic distinctions that are relevant to speech categories (e.g. Kingston & Diehl 1994, Cho & Ladefoged 1999). The present findings corroborate the hypothesis that speakers indicate the grouping of their speech units by manipulating phonetic detail (Cho 2016), thereby highlighting the effects of prosody on segmental phonetics (i.e. the prosody-phonetics interface). Finally, this investigation presents further evidence that phonetic information relates to multiple levels of prosodic structure simultaneously.

Acknowledgements

I would like to thank Caroline L. Smith, and Taehong Cho for their insightful comments on an early version of this manuscript. I am also grateful to the anonymous reviewers for their suggestions and constructive criticism, all of which helped improve the chapter.

Appendix A English stimuli

A.1 /p/

(1) *petitions*
  a. You're talking about ordinary polls but it's not the same || petitions must be SIGNED to be valid
  b. Very often students do get to voice their concerns but these silly # petitions make no difference

(2) *patrolmen*
  a. Policemen can certainly arrest you but that's not the case || patrolmen only REINFORCE order
  b. I'm used to being stopped by the police but I have to say: those angry # patrolmen really scared me

A.2 /t/

(3) *tequila*
  a. I've checked labels plus I've tried both drinks, so I'm pretty certain || tequila is WEAKER than pure vodka

- b. I'm used to drinking strong liquor because I don't like beer but that nasty # tequila made me so sick

(4) *toccata(s)*

- a. No, they're not something you eat at all! I know from music class || toccatas are just long MUSIC pieces
- b. I did enjoy it; she's an excellent musician, no doubt: that classy # toccata was fantastic

A.3 /k/

(5) *katrina*

- a. A lot of hurricanes do hit those parts but this time you are wrong || Katrina hit NEW ORLEANS, not Texas
- b. I've lived through many horrible storms that caused much damage but that deadly # Katrina destroyed the land

(6) *capacious*

- a. It doesn't refer to ability! You can check for yourself || capacious means ROOMY or full of space
- b. It is very sad there's not too much they can do at this point: the city's # capacious museum closed

Appendix B Spanish stimuli

B.1 /p/

(7) *pitada* [piˈta.ða]

- a. Estás confundido || pitada quiere decir SOPLADO más que sonido o pitido

  'You're mistaken, whistling is a BLOWING SOUND more than a noise or a beep'
- b. A causa de la lluvia, el árbitro Federico dio la # pitada a las tres horas

  'Because of the rain Federico the referee blew the whistle to end the match at 3 o'clock'

(8) *patrulla* [paˈtɾuja]

- a. Aqui en México || patrulla quiere decir un CARRO de vigilancia en la ciudad

  'Here in Mexico, a patrol is a CAR used by city police'

b. A pesar de las protestas, el gobierno va a mantener la # patrulla policial diaria

  ‘Despite the demonstrations, the government is keeping the daily police patrols’

B.2 /t/

(9) *tipazo* [tiˈpa.so]

a. No te confundas || tipazo quiere decir AMABLE más que un cuerpo atractivo

  ‘Don’t mix the two up, a stud is more like a NICE guy than a hot one’

b. Es un tanto vulgar, aquí en esta zona no se dice # tipazo a las personas

  ‘That’s a little vulgar; around here we don’t call anyone a stud’

(10) *tacada* [taˈka.ða]

a. Según sus abuelos || tacada tiene que ver con ARMAS de fuego y no con el billar

  ‘According to his grandparents, a strike is something to do with GUNS, not with playing pool’

b. Ganó el partido porque su papá le enseñó una # tacada spectacular

  ‘S/he won the match because her/his dad taught her/him a great move’

B.3 /k/

(11) *cuchara* [kuˈtʃaɾa]

a. Aprendí con ellos || cuchara se refiere TAMBIÉN a la herramienta del albañil

  ‘I learned this from them “cuchara” ALSO means a trowel that you use to build stuff’

b. Los albañiles estuvieron varias horas buscando la # cuchara para el muro

  ‘The contractors spent several hours looking for a trowel to build the wall’

(12) *capricho* [kaˈpɾitʃo]

a. Me parece raro || capricho significa un DESEO irracional muy intenso

  ‘That sounds strange, a whim means an irrational DESIRE that is very intense’

b. Los abuelos prepararon recetas para cumplirle su # capricho gastronómico
‘Her/His grandparents cook recipes just to satisfy her/his food whims’

Appendix C Portuguese stimuli

C.1 /p/

(13) *pitada* [pɪˈta.də]

a. Tem muito sal aqui || pitada quer dizer só UM POUQUINHO do ingrediente na receita
‘You put too much salt in this; a pinch means JUST A LITTLE of the ingredient’

b. Pimenta caiena é mais forte do que do reino; só uma # pitada tá mais que bom
‘Cayenne pepper is much stronger than black pepper; just a pinch is more than enough’

(14) *patola* [paˈtɔ.lə]

a. Não é um pato não || patola tem a ver com TAMANHO ou peso duma pessoa
‘It doesn’t mean full of stock, stocky has to do with someone’s SIZE or weight’

b. Elas venderam todos os filhotes mas essa cachorrinha # patola ninguém levou
‘They sold most of the puppies but no one really wanted to take the stocky one’

C.2 /t/

(15) *tutela* [tʊˈtɛ.lə]

a. Isso é outra coisa || tutela garante a AUTORIDADE sobre uma criança
‘That’s something else entirely; guardianship means having LEGAL AUTHORITY over a child’

b. Meu pai ficou sabendo outro dia que o Gilberto perdeu a # tutela dos três filhos
‘My father heard the other day that Gilberto lost custody of his three children’

(16) *tacada* [taˈka.də]

a. Esquece de taco || tacada quer dizer uma IDEIA inteligente que deu certo

‘Forget about the word taco; tacada means a clever IDEA that panned out’

b. Mesmo sem conhecer o gerente, não dá pra negar que aquela # tacada foi de mestre

‘You don’t have to know the manager to acknowledge that his clever move was exceptional’

C.3 /k/

(17) *cutelo* [kʊˈtɛ.lʊ]

a. Não é de açougue || cutelo é meio que um facão PEQUENO de uso diário

‘It’s not a butcher knife, a cleaver is a kind of SMALL hatchet for daily use in the kitchen’

b. Dependendo do tipo de carne é melhor usar aquele # cutelo maiorzinho

‘I guess it depends on the kind of meat but you should probably use that largish cleaver over there’

(18) *capela* [kaˈpɛ.lə]

a. Não é igrejinha || capela é um nicho PEQUENO dedicado a algum santo

‘It’s not a small church, a chapel is a small area dedicated to a given Catholic saint’

b. De todas as partes da igreja a que eu mais gosto é aquela # capela dourada lá

‘Of all the areas of the church, my favorite spot is that golden chapel over there’

Appendix D Statistical summaries

Table 11: Statistics summary for the English data (ms: milliseconds; dur: duration; m̃: median; MAD: median absolute deviation).

	Prosodic context							
	IP-initial				IP-medial			
Variables	min	max	m̃	MAD[a]	min	max	m̃	MAD
VOT /p t k/ (ms)	8	84	36	12	7	83	32	13
Vowel dur (ms)	15	73	32	12	21	71	33	12
Vowel F1 (normalized)[b]	7.73	22.4	9.95	1.33	6.23	26.54	9.96	1.81
Dur pause (ms)	32	1187	101	148	0	48	1	0
Dur stressed syllable (ms)	102	329	201	40	131	346	210	39

---

[a] MAD: median absolute deviation. MAD is a more robust measure of variability in non-normal distributions than the standard deviation (Levshina 2015).

[b] Obtained for each token based on the means and standard deviations calculated over all productions by the same speaker + 10.

Table 12: Statistics summary for the Spanish data (ms: milliseconds; dur: duration; m̃: median; MAD: median absolute deviation).

	Prosodic context							
	IP-initial				IP-medial			
Variables	min	max	m̃	MAD	min	max	m̃	MAD
VOT /p t k/ (ms)	4	62	15	6	4	46	17	7
Vowel dur (ms)	23	99	59	16	24	98	54	14
Vowel F1 (normalized)	8.14	13.87	9.86	1.36	8.36	13.15	9.56	1.16
Dur pause (ms)	59	1164	125	131	0	89	3	0
Dur stressed syllable (ms)	109	375	201	31	142	364	212	30

Table 13: Statistics summary for the Portuguese data (ms: milliseconds; dur: duration; MAD: median absolute deviation).

	Prosodic context							
	IP-initial				IP-medial			
Variables	min	max	m̃	MAD	min	max	m̃	MAD
VOT /p t k/ (ms)	7	58	20	12	1	51	22	13
Vowel dur (ms)	19	117	56	21	16	100	45	21
Vowel F1 (normalized)	8.07	12.27	9.95	1.02	8.20	12.96	10.21	1.26
Dur pause (ms)	47	724	185	152	0	33	0	0
Dur stressed syllable (ms)	120	468	246	71	133	385	209	37

Appendix E Wilcoxon tests results

Table 14: Results of two-tailed Wilcoxon tests for statistically significant differences between experimental conditions (IP-initial vs. IP-medial position).

	English		Spanish		Portuguese	
Variable	*W*	*p*	*W*	*p*	*W*	*p*
VOT /p t k/	26472	0.015*	19958	0.265	22193	0.917
Vowel duration	27261	0.487	25630	<0.001***	30646	<0.001***
Vowel F1	22022	0.917	22846	0.178	20057	0.098
Dur pause	51859	<0.001***	40296	<0.001***	42210	<0.001***
Dur stressed syllable	23184	<0.001***	18090	<0.001***	30630	<0.001***

References

Abramson, Arthur S. & Douglas H. Whalen. 2017. Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63. 75–86. DOI: 10.1016/j.wocn.2017.05.002.

Ahn, Suzy. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics* 71. 451–467. DOI: 10.1016/j.wocn.2018.10.003.

Alves, Mariane Antero, Izabel Christine Seara, Fernando Santana Pacheco, Simone Klein & Rui Seara. 2008. On the voiceless aspirated stops in Brazilian Portuguese. In *Computational processing of the Portuguese language (PROPOR 2008)*, 248–251. Springer. DOI: 10.1007/978-3-540-85980-2_31.

Audacity Team. 2014. *Audacity: Free audio editor and recorder.* https://www.audacityteam.org/.

Avelino, Heriberto. 2018. Mexico City Spanish. *Journal of the International Phonetic Association* 48(2). 223–230. DOI: 10.1017/S0025100316000232.

Barnes, Jonathan A. 2006. *Strength and weakness at the interface: Positional neutralization in phonetics and phonology.* Berlin/NewYork: Mouton de Gruyter. DOI: 10.1515/9783110197617.

Beckman, Mary E. 1986. *Stress and non-stress accent.* Dordrecht: Foris Publications. DOI: 10.1515/9783110874020.

Beckman, Mary E., Kenneth J. de Jong, Sun-Ah Jun & Sook-Hyang Lee. 1992. The interaction of coarticulation and prosody in sound change. *Language and Speech* 35(1). 45–58. DOI: 10.1177/002383099203500205.

Beckman, Mary E. & Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In *Papers in Laboratory Phonology III: Phonological structure and phonetic form*, 7–33. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511659461.002.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309. DOI: 10.1017/s095267570000066x.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111. DOI: 10.1016/j.jml.2008.06.003.

Beňuš, Štefan & Juraj Šimko. 2014. Emergence of prosodic boundary: Continuous effects of temporal affordance on inter-gestural timing. *Journal of Phonetics* 44. 110–129. DOI: 10.1016/j.wocn.2013.12.005.

Berkovits, Rochele. 1994. Durational effects in final lengthening, gapping and contrastive stress. *Language and Speech* 37. 237–250. DOI: 10.1177/002383099403700302.

Bodur, Kübra, Sweeney Branje-Canioni, Morgane Peirolo, Ingrid Tiscareno & James German. 2021. Domain-initial strengthening in Turkish: Acoustic cues to prosodic hierarchy in stop consonants. In *Proceedings of InterSpeech*, 1459–1463. DOI: 10.21437/interspeech.2021-2230.

Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer.* http://www.praat.org/.

Bombien, Lasse, Christine Mooshammer, Phil Hoole & Barbara Kühnert. 2010. Prosodic and segmental effects on EPG contact patterns of word-initial German clusters. *Journal of Phonetics* 38(3). 388–403. DOI: 10.1016/j.wocn.2010.03.003.

Byrd, Dani, Jelena Krivokapić & Sungbok Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the Acoustical Society of America* 120. 1589–1599. DOI: 10.1121/1.2217135.

Byrd, Dani & Daylen Riggs. 2008. Locality interactions with prominence in determining the scope of phrasal lengthening. *Journal of the International Phonetic Association* 38(2). 187–202. DOI: 10.1017/s0025100308003460.

Câmara JR, Joaquim Mattoso. 1972. *The Portuguese language*. Chicago: University of Chicago Press.

Cambier-Langeveld, Tina. 1997. The domain of final lengthening in the production of Dutch. In Jane A. Coerts & Helen de Hoop (eds.), *Linguistics in the Netherlands*, 13–24. Amsterdam: Benjamins. DOI: 10.1075/avt.14.04cam.

Cantoni, Maria Mendes. 2013. *O acento no português brasileiro: uma abordagem experimental [Accent in Brazilian Portuguese: An experimental approach]*. Belo Horizonte, Brazil: Universidade Federal de Minas Gerais. (Doctoral dissertation). http://hdl.handle.net/1843/LETR-96NLVH.

Carlson, Katy. 2009. How prosody influences sentence comprehension. *Language and Linguistics Compass* 3(5). 1188–1200. DOI: 10.1111/j.1749-818x.2009.00150.x.

Cho, Taehong. 2011. Laboratory phonology. In Nancy C. Kula, Bert Botma & Kuniya Nasukawa (eds.), *The Continuum Companion to phonology*, 343–368. London: Continuum.

Cho, Taehong. 2015. Language effects on timing at the segmental and suprasegmental levels. In Melissa A. Redford (ed.), *The handbook of speech production*, 505–559. Chichester: Wiley-Blackwell. DOI: 10.1002/9781118584156.ch22.

Cho, Taehong. 2016. Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass* 10(3). 120–141. DOI: 10.1111/lnc3.12178.

Cho, Taehong & Patricia A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29(2). 155–190. DOI: 10.1006/jpho.2001.0131.

Cho, Taehong & Patricia A. Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37(4). 466–485. DOI: 10.1016/j.wocn.2009.08.001.

Cho, Taehong, Daejin Kim & Sahyang Kim. 2017. Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics* 64. 71–89. DOI: 10.1016/j.wocn.2016.12.003.

Cho, Taehong, Jiseung Kim & Sahyang Kim. 2013. Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *The Journal of the Acoustical Society of America* 133. 384–390. DOI: 10.1121/1.4800179.

Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27. 207–229. DOI: 10.1006/jpho.1999.0094.

Cho, Taehong & James M. McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33(2). 121–157. DOI: 10.1016/j.wocn.2005.01.001.

Cho, Taehong, James M. McQueen & Ethan A. Cox. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35(2). 210–243. DOI: 10.1016/j.wocn.2006.03.003.

Cho, Taehong, Douglas H. Whalen & Gerard Docherty. 2019. Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics* 72. 52–65. DOI: 10.1016/j.wocn.2018.11.002.

Clopper, Cynthia G. & Rajka Smiljanic. 2011. Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics* 39(2). 160. DOI: 10.1016/j.wocn.2011.02.006.

Cole, Jennifer. 2015. Prosody in context: A review. *Language, Cognition and Neuroscience* 30(1-2). 1–31. DOI: 10.1080/23273798.2014.963130.

Cole, Jennifer, Heejin Kim, Hansook Choi & Mark Hasegawa-Johnson. 2007. Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* 35(2). 180–209. DOI: 10.1016/j.wocn.2006.03.004.

Cristófaro-Silva, Thaïs, Izabel Christine Seara, Adelaide Silva, Andreia Schurt Rauber & Maria Cantoni. 2019. *Fonética acústica: Os sons do Português brasileiro [Acoustic phonetics: The sounds of Brazilian Portuguese]*. São Paulo: Contexto.

Crosswhite, Katherine M. 2001. *Vowel reduction in Optimality Theory*. New York: Routledge.

Crystal, Thomas H. & Arthur S. House. 1988. The duration of American-English vowels: An overview. *Journal of Phonetics* 16(3). 263–284. DOI: 10.1016/s0095-4470(19)30500-5.

Davidson, Lisa. 2006. Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica* 63(2-3). 79–112. DOI: 10.1159/000095304.

Davies, Mark. 2008. The corpus of contemporary American English (COCA): 560 million words. https://www.english-corpora.org/coca/.

Davies, Mark. 2016. Corpus del Espanol: Web/dialects. https://www.corpusdelespanol.org.

Davies, Mark & Michael Ferreira. 2006. Corpus do Portugues. https://www.corpusdoportugues.org.

de la Mota, Carme, Pedro M. Butragueño & Pilar Prieto. 2010. Mexican Spanish intonation. In Pilar Prieto & Paolo Roseano (eds.), *Transcription of intonation of the Spanish language*, 319–350. Munich: LINCOM Europa. http://prosodia.upf.edu/home/arxiu/activitats/4th_workshop/protegit/Mexican_final.pdf.

de Jong, Kenneth J. 2004. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics* 32(4). 493–516. DOI: 10.1016/j.wocn.2004.05.002.

de Jong, Nivja H. & Ton Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41(2). 385–390. DOI: 10.3758/brm.41.2.385.

Delforge, Ann Marie. 2008. Gestural alignment constraints and unstressed vowel devoicing in Andean Spanish. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 147–155. http://www.lingref.com/cpp/wccfl/26/paper1668.pdf.

Dilley, Laura, Stefanie Shattuck-Hufnagel & Mari Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24(4). 423–444. DOI: 10.1006/jpho.1996.0023.

Erickson, Donna. 1998. Effects of contrastive emphasis on jaw opening. *Phonetica* 55(3). 147–169. DOI: 10.1159/000028429.

Fletcher, Janet. 2010. The prosody of speech: Timing and rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The handbook of phonetic sciences*, 521–602. Chichester: Wiley. DOI: 10.1002/9781444317251.ch15.

Fónagy, Ivan & Klara Magdics. 1960. Speed of utterance in phrases of different lengths. *Language and Speech* 3(4). 179–192. DOI: 10.1177/002383096000300401.

Fougeron, Cecile. 1999. Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics* 97. 1–73. https://escholarship.org/uc/item/3pg0w66x.

Fougeron, Cecile. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29(2). 109–135. DOI: 10.1006/jpho.2000.0114.

Fougeron, Cecile & Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101(6). 3728–3740. DOI: 10.1121/1.418332.

Fowler, Carol A. & Jonathan Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26(5). 489–504. DOI: 10.1016/0749-596X(87)90136-7.

Frota, Sonia. 2000. *Prosody and focus in European Portuguese: Phonological phrasing and intonation*. New York/London: Garland Pub. DOI: 10.4324/9781315054384.

Frota, Sonia, Marisa Cruz, Flaviane Fernandes-Svartman, Gisela Collischonn, Aine Fonseca, Carolina Serra, Pedro Oliveira & Marina Vigário. 2015. Intonational variation in Portuguese: European and Brazilian varieties. In Sonia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 235–283. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199685332.003.0007.

Frota, Sonia, Mariapola D'Imperio, Gorka Elordieta, Pilar Prieto & Marina Vigário. 2007. The phonetics and phonology of intonational phrasing in Romance. In *Segmental and prosodic issues in Romance phonology*, 131–152. Benjamins. DOI: 10.1075/cilt.282.10fro.

Frota, Sonia & Joao Moraes. 2016. Intonation of European and Brazilian Portuguese. In W. Leo Wetzels, Joao Costa & Sergio Menuzzi (eds.), *The handbook of Portuguese linguistics*, 141–166. Cambridge, MA: Wiley-Blackwell. DOI: 10.1002/9781118791844.ch9.

Frota, Sonia & Pilar Prieto. 2015. *Intonation in Romance*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199685332.001.0001.

Garellek, Marc. 2015. Perception of glottalization and phrase-final creak. *The Journal of the Acoustical Society of America* 137(2). 822–831. DOI: 10.1121/1.4906155.

Georgeton, Laurianne & Cecile Fougeron. 2014. Domain-initial strengthening on French vowels and phonological contrasts: Evidence from lip articulation and spectral variation. *Journal of Phonetics* 44. 83–95. DOI: 10.1016/j.wocn.2014.02.006.

Goldman, Jean-Philippe. 2011. EasyAlign: An automatic phonetic alignment tool under Praat. In *Proceedings of InterSpeech*. DOI: 10.21437/interspeech.2011-815.

Gries, Stefan Thomas. 2013. *Statistics for linguistics with R: A practical introduction*. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110307474.

Gussenhoven, Carlos & Antonius C. M. Rietveld. 1992. Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics* 20. 283–303. DOI: 10.1016/s0095-4470(19)30636-9.

Hsu, Chai-Shune & Sun-Ah Jun. 1998. Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics* 96. 69–89.

Hualde, José Ignacio. 2007. Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics* 5(6). 59–89. DOI: 10.5334/jpl.145.

Hualde, José Ignacio. 2012. Stress and rhythm. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The handbook of Hispanic linguistics*, 153–171. Chichester: Wiley-Blackwell. DOI: 10.1002/9781118228098.ch8.

Hualde, José Ignacio & Marianna Nadeu. 2014. Rhetorical stress in Spanish. In Harry van der Hulst (ed.), *Word stress: Theoretical and typological issues*, 228–252. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9781139600408.010.

Hualde, José Ignacio & Pilar Prieto. 2015. Intonational variation in Spanish: European and American varieties. In Sonia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 350–392. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199685332.003.0010.

Jun, Sun-Ah. 2005. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.001.0001.

Jun, Sun-Ah. 2006. Intonation. In Lynn Nadel (ed.), *Encyclopedia of cognitive science*, 618–624. Chichester: Wiley-Blackwell. DOI: 10.1002/0470018860.s00241.

Jun, Sun-Ah. 2014. *Prosodic typology II: The phonology of intonation and phrasing*. Sun-Ah Jun (ed.). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199567300.001.0001.

Jun, Sun-Ah & Mary E. Beckman. 1994. Distribution of devoiced high vowels in Korean. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, 479–482. DOI: 10.21437/icslp.1994-110.

Katsika, Argyro. 2016. The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics* 55. 149–181. DOI: 10.1016/j.wocn.2015.12.003.

Keating, Patricia A. 2004. Phonetic encoding of prosodic structure. *UCLA Working Papers in Phonetics* 103. 48–63. http://phonetics.linguistics.ucla.edu/workpapph/103/4-Keating-ISSPSydney.pdf.

Keating, Patricia A., Taehong Cho, Cecile Fougeron & Chai-Shune Hsu. 2003. Domain-initial strengthening in four languages. In *Papers in Laboratory Phonology VI: Phonetic interpretation*, 145–163. Cambridge: Cambridge University Press.

Kessinger, Rachel H. & Sheila E. Blumstein. 1998. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 26(2). 117–128. DOI: 10.1006/jpho.1997.0069.

Kim, Sahyang, Jiseung Kim & Taehong Cho. 2018. Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics* 71. 65–80. DOI: 10.1016/j.wocn.2018.07.004.

Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70(3). 419–454. DOI: 10.2307/416481.

Krivokapić, Jelena. 2014. Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1658). 1–10. DOI: 10.1098/rstb.2013.0397.

Krull, Diana. 1997. Prepausal lengthening in Estonian: Evidence from conversational speech. In Ilse Lehiste & Jaan Ross (eds.), *Estonian prosody: Papers from a symposium*, 136–148.

Kuzla, Claudia, Taehong Cho & Mirjam Ernestus. 2007. Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics* 35(3). 301–320. DOI: 10.1016/j.wocn.2006.11.001.

Ladd, D. Robert. 2008. *Intonational phonology*. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511808814.

Lavoie, Lisa M. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Garland. 254. DOI: 10.4324/9780203826423.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: Benjamins. DOI: 10.1075/z.195.

Liberman, Mark Y. & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2). 249–336. https://www.jstor.org/stable/4177987.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *WORD* 20(3). 384–422. DOI: 10.1080/00437956.1964.11659830.

Mateus, Maria Helena & Ernesto D'Andrade. 2000. *The phonology of Portuguese*. Oxford: Oxford University Press.

Mo, Yoonsook. 2007. Temporal, spectral evidence of devoiced vowels in Korean. In *Proceedings of the 16th ICPhS*, 445–448. http://www.icphs2007.de/conference/Papers/1597/1597.pdf.

Nadeu, Marianna. 2014. Stress- and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics* 46. 1–22. DOI: 10.1016/j.wocn.2014.05.003.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris Publications. DOI: 10.1515/9783110977790.

Oh, Sejin. 2021. Phonetic and phonological vowel reduction in Brazilian Portuguese. *Phonetica*. 435–465. DOI: 10.1515/phon-2021-2012.

Oller, D. Kimbrough. 1973. The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America* 54(5). 1235–1247. DOI: 10.1121/1.1914393.

Onaka, Akiko, Sallyanne Palethorpe, Catherine Watson & Jonathan Harrington. 2003. An acoustic analysis of domain-initial strengthening effect in Japanese. In *Proceedings of the sixth international seminar on speech production*, 201–206.

Ortega-Llebaria, Marta & Pilar Prieto. 2007. Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In *Segmental and prosodic issues in Romance phonology*, 155–175. Amsterdam: Benjamins. DOI: 10.1075/cilt.282.11ort.

Parrell, Benjamin. 2014. *Dynamics of consonant reduction*. University of Southern California. (Doctoral dissertation).

Petrone, Caterina, Mariapola D'Imperio, Leonardo Lancia & Susanne Fuchs. 2014. The interplay between prosodic phrasing and accentual prominence on articulatory lengthening in Italian. In *Proceedings of Speech Prosody*, 192–196. DOI: 10.21437/speechprosody.2014-26.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. Massachusetts Institute of Technology. (Doctoral dissertation). http://dspace.mit.edu/handle/1721.1/16065.

Pierrehumbert, Janet B. & Julia Bell Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 271–312. Cambridge, MA: MIT Press. DOI: 10.7916/D8KD24FP.

Pierrehumbert, Janet B. & David Talkin. 1992. Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II: Gesture, segment, prosody*, 90–117. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511519918.005.

Plag, Ingo, Gero Kunter & Mareile Schramm. 2011. Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics* 39(3). 362–374. DOI: 10.1016/j.wocn.2011.03.004.

Prieto, Pilar & Paolo Roseano. 2010. *Transcription of intonation of the Spanish language*. Munich: LINCOM Europa.

Prieto, Pilar, Jan van Santen & Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23(4). 429–51. DOI: 10.1006/jpho.1995.0032.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Ramanarayanan, Vikram, Erik Bresch, Dani Byrd, Louis Goldstein & Shrikanth S. Narayanan. 2009. Analysis of pausing behavior in spontaneous speech using real-time magneticresonance imaging of articulation. *The Journal of the Acoustical Society of America* 126(5). EL160–EL165. DOI: 10.1121/1.3213452.

Rao, Rajiv. 2010. Final lengthening and pause duration in three dialects of Spanish. In *Selected proceedings of the 4th conference on laboratory approaches to Spanish phonology*, 69–82. http://www.lingref.com/cpp/lasp/4/paper2368.pdf.

Redi, Laura & Stefanie Shattuck-Hufnagel. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29(4). 407–429. DOI: 10.1006/jpho.2001.0145.

Roettger, Timo & Matthew Gordon. 2017. Methodological issues in the study of word stress correlates. *Linguistics Vanguard* 3. 1. DOI: 10.1515/lingvan-2017-0006.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard & Jiahong Yuan. 2014. *FAVE: Forced alignment and vowel extraction).* Version 1.1.3. DOI: 10.5281/zenodo.9846.

Scheer, Tobias & Philippe Ségéral. 2008. Positional factors in lenition and fortition. In Joaquim Brandão de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition*, 131–172. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110211443.1.131.

Schubö, Fabian & Sabine Zerbian. 2023. The patterns of pre-boundary lengthening in German. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 1–34. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777526.

Selkirk, Elisabeth. 1984. *Prosody and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Shattuck-Hufnagel, Stefanie & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247. DOI: 10.1007/bf01708572.

Singmann, Henrik, Ben Bolker, Jake Westfall & Frederik Aust. 2015. *afex: Analysis of factorial experiments.* R package version 0.13–145. https://CRAN.R-project.org/package=afex.

Stevens, Kenneth N. & Samuel Jay Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65(1). 81–106. DOI: 10.2307/414843.

Tabain, Marija. 2003. Effects of prosodic boundary on /aC/ sequences: Articulatory results. *The Journal of the Acoustical Society of America* 113(5). 2834–2849. DOI: 10.1121/1.1564013.

Torreira, Francisco & Mirjam Ernestus. 2011. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology* 2(2). 331–353. DOI: 10.1515/labphon.2011.012.

Turk, Alice E., Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. In Stefan Sudhoff, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.), *Methods in empirical prosody research*, 1–28. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110914641.1.

Turk, Alice E. & James R. Sawusch. 1997. The domain of accentual lengthening in American English. *Journal of Phonetics* 25(1). 25–41. DOI: 10.1006/jpho.1996.0032.

Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28(4). 397–440. DOI: 10.1006/jpho.2000.0123.

Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35(4). 445–472. DOI: 10.1016/j.wocn.2006.12.001.

Vigário, Marina. 2003. *The prosodic word in European Portuguese*. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110900927.

Vogel, Irene, Angeliki Athanasopoulou & Natália Brambatti Guzzo. 2018. Timing properties of (Brazilian) Portuguese and (European) Spanish. In Lori Repetti & Francisco Ordóñez (eds.), *Romance languages and linguistic theory*, 325–340. Amsterdam: Benjamins. DOI: 10.1075/rllt.14.19vog.

White, Laurence. 2002. *English speech timing: A domain and locus approach*. University of Edinburgh. (Doctoral dissertation). https://www.cstr.ed.ac.uk/projects/eustace/fulltext/front.pdf.

White, Laurence, Silvia Benavides-Varela & Katalin Mády. 2020. Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? *Journal of Phonetics* 81. 10098. DOI: 10.1016/j.wocn.2020.100982.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America* 91(3). 1707–1717. DOI: 10.1121/1.402450.

Winitz, Harris, M. E. Scheib & James A. Reeds. 1972. Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *The Journal of the Acoustical Society of America* 51(4B). 1309–1317. DOI: 10.1121/1.1912976.

Xu, Yi. 2010. In defense of lab speech. *Journal of Phonetics* 38(3). 329–336. DOI: 10.1016/j.wocn.2010.04.003.

Chapter 3

Chunking an unfamiliar language: Results from a perception study of German listeners

Nele Ots[a] & Piia Taremaa[b]

[a]Goethe-University Frankfurt [b]University of Tartu

This study investigates the impact of prosodic boundary phenomena and syntactic clause boundaries on native and non-native speech chunking. German and Estonian listeners were asked to listen to spontaneous utterances spoken in Estonian and to mark in corresponding written transcripts when they perceived any sort of a break between the words. Estonian listeners were the strongest guided by the clause boundaries whereas German listeners were sensitive to all of the prosodic boundary phenomena but resistant to the presence of clause boundaries. In particular, both German and Estonian listeners utilized longer pauses and rising F0 contour as cues for chunk boundaries. German listeners additionally employed phrase-final lengthening and intensity drop. These results suggest strong bottom-up effects in non-native speech processing, and both bottom-up effects and top-down effects in native processing of speech. Thus, the well-known prosodic boundary phenomena trigger bottom-up processing in on-going spontaneous speech comprehension.

1 Introduction

Speech comprehension starts with and depends on the extraction of discrete sequential units from continuous speech flow. In order to discern and maintain these units in working memory, listeners interpret smaller units detected in the context of larger ones, that is, in (speech) chunks (Dahan & Ferreira 2019, Christiansen & Chater 2016). Speech chunking operates across multiple levels of

Nele Ots & Piia Taremaa. 2023. Chunking an unfamiliar language: Results from a perception study of German listeners. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 87–117. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777530

linguistic representation and is related to top-down as well as bottom-up processing (see, e.g., Dahan & Ferreira 2019). Top-down processing concerns world and linguistic knowledge (including lexical, semantic and syntactic knowledge), whereas bottom-up processing relates to sensory input from acoustic signals. Native listeners, when extracting discrete units from speech, are known to exploit their top-down knowledge about lexical-semantic and syntactic information (Mattys et al. 2005). Thus, top-down processing is decisive for chunking a continuous stream of speech into units. This speeds up spoken language comprehension by helping listeners to rapidly recognize and process segments of language that are syntactically, lexically and semantically coherent and plausible, given the context.

In certain aspects of linguistic structures, the extent of bottom-up processing in speech chunking is unclear. In particular, signal-driven prosodic cues (e.g., lengthening of the segments) have proven to be highly functional for the recognition of words (e.g., White et al. 2020). Whether listeners are also able to recognize chunks at a higher level, i.e., intonational phrases, is only vaguely understood. Recently, Ordin et al. (2017) pointed out that listeners may apply phrase-level prosody alongside word-level prosody for the generation of so-called prosodic frames (Keating & Shattuck-Hufnagel 2002, Schild et al. 2014, Silbert et al. 2014). Understanding the role of phrase-level prosody in chunking processes is necessary because prosodic frames are proposed to take part in encoding as well as decoding processes in language production and perception (Keating & Shattuck-Hufnagel 2002, Schild et al. 2014, Silbert et al. 2014).

A handful of phonetic perception studies have indicated that signal-driven prosodic information (e.g., durational or tonal discontinuities in terms of pausing, lengthening and pitch reset) ranks rather low in native speech chunking (Cole et al. 2010, Christodoulides et al. 2018, Duez 1985). In contrast, several psycholinguistic studies have demonstrated the significant role of phrasal prosody in recognizing and remembering novel words (Langus et al. 2012, Ordin et al. 2017). This, in turn, drives our investigation of speech chunking in non-native listeners in comparison with native listeners (for a similar approach, see Himmelmann et al. 2018, Riesberg et al. 2020). The current study asks to what extent speech chunks are accessible from signal-driven prosodic information.

1.1 Signal-driven prosodic boundary cues

A type of well-known prosodic unit is the so-called tone group (Halliday 1967), also known as the intonation unit (Chafe 1987) or, more commonly, an intonational phrase (Pierrehumbert 1980, Ladd 2008). In intonational phonology, intonational phrases (IPs) constitute abstract phonological units that are composed of discrete abstract categories of pitch accents and boundary tones (Ladd 2008, Pierrehumbert & Hirschberg 1990). The identification of phonological categories of pitch accents and boundary tones in spoken sentences usually follows from phonological analysis. As such, the well-known concept of IP constitutes top-down information about the phonological structure of a language. For the purposes of this study, it is interesting that IPs frequently correspond with concrete acoustic regularities directly observable in the speech signal.

IPs are frequently characterized as units of tonal coherence (Bois et al. 1992, Breen et al. 2012, Buhmann et al. 2002, Himmelmann et al. 2018). An underlying acoustical phenomenon of the significant percept of tonal coherence is the continuous decline of fundamental frequency (F0, acoustic approximation of sentence intonation) from the beginning to the end of an IP. This decline was traditionally measured considering the F0 maxima in phonological pitch accents (see, e.g., Ladd 1988, Liberman & Pierrehumbert 1984, Pierrehumbert 1979). More recent research has found a more automatic way to fit a straight line to the F0 contour as a function of time (Yuan & Liberman 2014). F0 declination is a global component of the F0 contour, and it should interact only mildly with local F0 movements determining pitch accents and boundary tones (Fujisaki 1983, Fujisaki & Hirose 1982). The regularity of F0 declination underlies the readily audible cue of pitch reset, which means that the continuous decline of the F0 contour is disrupted by setting the level of F0 much higher than predicted by an on-going F0 decline, e.g., by stepping up the pitch. In intonation research, the pitch reset has often been utilized as a valuable cue signaling the right edge of an IP (Cooper & Sorensen 1981, Couper-Kuhlen 2001, Himmelmann et al. 2018, Ladd 1988, Schuetze-Coburn et al. 1991, Thorsen 1985).

In the stream of speech, IPs are most easily defined by their boundaries. The IP boundaries in spoken language are associated with a battery of phonetic boundary phenomena encompassing systematic changes in duration, intensity and F0. In particular, a durational discontinuity that involves slowing down speech, or, more specifically, lengthening speech segments, constitutes a type of signal-driven prosodic cue that is frequently referred to as pre-boundary or phrase-final lengthening (Berkovits 1994, Fon et al. 2011, Nakai et al. 2009, Petrone et al. 2017, Wightman et al. 1992; for this cue in German, see also Schubö & Zerbian 2023 [this volume], Huttenlauch et al. 2023 [this volume], and Wellmann et al. 2023 [this volume]). In terms of prosodic boundary cues, intensity has attracted interest to a lesser degree. However, some studies have indicated that an intensity curve within words may also function as a boundary cue. The increasing intensity difference between the initial and final syllable in a word constitutes the phenomenon of intensity drop (Trouvain et al. 1998, Wagner & McAuliffe 2019). Finally, IP boundaries are well known to be indexed by intonational movements (boundary tones at the abstract level of intonational phonology), which are indexed by falling or rising F0 contours at the ends of IPs (e.g., consider falling intonation in statements and rising intonation in questions; on the role of rising F0 contours, see Petrone et al. 2017, Huttenlauch et al. 2023 [this volume], and Wellmann et al. 2023 [this volume]). Thus, signal-driven prosodic cues, such as phrase-final lengthening, intensity drop, pitch reset and F0 movements, define IPs in spoken utterances. These cues have certain acoustic correlates in the speech signal and these can be investigated as input for bottom-up processing.

1.2 Perception of phrasal prosody

In the auditory processing of language, the discontinuities of duration and F0 have been shown to be highly functional. Namely, phrase-final lengthening and continuous F0 declination can help listeners to discover long-distance dependencies between words, or tentatively, clausal relationships (de la Cruz-Pavía et al. 2019, Langus et al. 2012, Ordin et al. 2017). For example, in an experiment with Italian listeners, Langus et al. (2012) created a novel language by defining words and long-distance semantic dependencies between them through systematically manipulating the probability distributions of sounds and syllables. Importantly, the stipulative sentences of the novel language were additionally accompanied by pre-boundary lengthening and continuous F0 declination. Langus et al. (2012) were able to demonstrate that long-distance dependencies between the words were only discovered in the presence of prosodic cues. Moreover, they found that while F0 declination is useful for detecting dependencies at the level of a stipulative sentence or a clause, phrase-final lengthening induces the listener to perceive a stipulative syntactic phrase. Thus, they were able to separate the functions of the two types of prosodic cues at two different linguistic levels – a stipulative phrase vs. a stipulative clause. Altogether, the results from Langus et al. (2012) demonstrate that the presence of phrase-final lengthening and F0 declination clearly enforces perception of a sort of language chunk. For additional functionality in infant language acquisition see Wellmann et al. 2023 [this volume].

Phonetic studies of perceptual speech chunking further indicate that there is an unavoidable syntactic component in the perceptual chunking of language. For example, Duez (1985) presented listeners with natural, distorted and synthesized speech and asked them to explicitly mark silent pauses. Remarkably, the results show that listeners detected significantly fewer pauses in distorted and synthesized speech than in normal speech. Thus, the study indicates that listeners, even when explicitly detecting signal-based information, may rely more strongly on syntactic-semantic information than acoustic information or may ignore the latter altogether (for a replication of these results, see Simon & Christodoulides 2016). In a more recent study, Cole et al. (2010) asked native listeners of American English to listen to broadcasted conversations and to mark in written transcripts where they heard some sort of a break or juncture. The results show that clause boundaries had the strongest impact on boundary perception; phrase-final lengthening or a duration cue ranked lower, and F0 did not play any role. A study by Christodoulides et al. (2018) employed slightly different methodology by asking French listeners to press a button when they heard some sort of a break in speech. The timeline of button presses was synchronized with the stream of speech, and without having any written input, the outcome was nevertheless that syntactic clause boundaries most strongly contributed to boundary perception. These results further demonstrate how influential the access to clausal information is in the metalinguistic tasks. As the clause boundaries constitute the linguistic knowledge, they can be taken as input for the top-down processing. As such, the existing studies demonstrate pervasive top-down processing in native speech comprehension.

Strikingly, Riesberg et al. (2020) found that lexical and syntactic variables participate even in non-native perception of speech chunks. Their study employed the same methodology as in Cole et al. (2010) and asked native speakers of German to listen to short stories spoken in Papuan Malay and mark in written transcripts where they heard some sort of a break. Speakers of Papuan Malay were presented with short stories in German with the same task. Both language groups also judged the stories spoken in their native language. For listeners from both language groups, clause boundaries were the second strongest factor that contributed to the perception of chunks in the unfamiliar languages, while pauses were the strongest cue.

This result becomes less surprising when considering language production. Specifically, several studies have found that syntactically defined segments, such as clauses and phrases, are often accompanied by acoustic discontinuities (see, e.g., Cutler et al. 1997, Petrone et al. 2017). For example, Féry & Ishihara (2009) demonstrated in a reading experiment that speakers tended to reset pitch and start a new declination trend for F0 at the beginning of embedded subclauses. This indicates that IPs, or the signal-driven prosodic cues of duration, and F0 in particular, tend to strongly associate with the syntactic representation of language. In other words, syntactic elements such as clauses are produced as prosodically coherent speech chunks, and the results from Riesberg et al. (2020) suggest they will also be perceived as such regardless of the listener's language background.

1.3 The current study

The aim of this study is to determine the impact of signal-driven prosodic cues (i.e., bottom-up processing) separately from syntactic-semantic information (i.e., top-down processing) in speech chunking. For this, we investigate how non-native speakers perceive an unfamiliar natural language. When processing an unfamiliar language, semantic-syntactic cues are not available to the listener. Arguably, this forces non-native listeners to rely on signal-based acoustic cues whilst chunking speech (Himmelmann et al. 2018, Riesberg et al. 2020). By investigating chunking of speech flow in an unfamiliar language, we are able to examine the role of signal-based prosodic information in bottom-up processing of language.

To assess the influence of prosodic information on speech chunking, we conducted a chunking experiment based on Rapid Prosody Transcription (RPT; Cole et al. 2010, 2011, Mahrt 2016) in which Estonian and German listeners had to chunk excerpts of spontaneous utterances spoken in Estonian. We investigated the impact of signal-driven prosodic cues (i.e., phrase-final lengthening, intensity drop, rate of F0 declination, and pause duration) against the clausal structure of spontaneously spoken utterances and the listeners' language background. Signal-driven prosodic information serves as input for bottom-up processing, whereas the clausal structure provides input for top-down processing. Crucially, the clausal structure of Estonian utterances is not available for German listeners who are unfamiliar with the Estonian language. Therefore, we hypothesized that the impact of bottom-up information in speech chunking is modulated by the listener's language background. Based on the notion of top-down processing, we expected the German listeners to be less affected by the clausal structure and to be more sensitive to the signal-driven prosodic cues, whereas the Estonian listeners were expected to use both clausal and acoustic cues. The alternative prediction relies on the results in Riesberg et al. (2020). Namely, the German listeners could perform similarly to Estonian listeners in terms of clausal cues. This outcome would indicate a strong relationship between prosodic information and clausal structure in Estonian speech production because, arguably, the German listeners would rely on the prosodic cues that are tightly associated with clausal structure.

2 Materials and method

For our experiment, we applied the methodology of Rapid Prosody Transcription (RPT), in which listeners are typically asked to listen to excerpts of speech and mark the words that they perceive as prominent or that stand before some sort of a break (Cole et al. 2010, 2011, Mahrt 2016).

2.1 Participants

Altogether, 47 Estonian listeners (average age 30.0 years) took part in an earlier experiment (Ots & Taremaa 2022). They originated from various regions of Estonia. Given their age, they most likely speak Standard Estonian, and the dialectal variation in Estonia is probably not that pronounced in young speakers.

For this study, 90 native speakers of German were recruited through a crowdsourcing marketplace designed for conducting research (Prolific). They were paid about £2.50 to complete the task, which took about 20 minutes. The average age of the participants was 28.8 years (with 0.03 percent of participants not reporting). 48.9 percent of participants were female, and 46.7 percent were male (with 0.04 percent of participants not reporting). All participants reported German to be their first language. 86.7 percent of participants reported having knowledge of some other language, most frequently English. None reported having knowledge of Estonian.

2.2 Stimuli

We extracted 396 excerpts of spontaneous speech (4727 words altogether) from 10 native Estonian speakers (5 male and 5 female speakers with an average age of 25.3 years) from the phonetic corpus of spoken Estonian (Lippus et al. 2016). Auditive analysis did not reveal any distinctive dialectal characteristics in these speakers. They appeared to use Standard Estonian as it is taught in schools. The excerpts constituted a stretch of fluent speech between silent pauses of 400 ms or longer. The excerpts contained 18 to 24 syllables, yielding an average duration of 3300 ms. For the experiment with Estonian listeners, the 396 excerpts were randomly distributed between 4 different lists, each containing 99 excerpts in total. The lists for German listeners were kept shorter, as their task was to listen to non-native language. Thus, the 396 excerpts were randomly distributed between 9 lists, with each list containing 44 excerpts in total.

2.3 Procedure

The Estonian excerpts were presented to native speakers of German, unfamiliar with the Estonian language. The study was conducted over the internet using LMEDS software (Mahrt 2016). Based on RPT methodology (see, e.g., Cole et al. 2010, 2011, Riesberg et al. 2020), the participants were asked to listen to speech excerpts and identify the chunks of words ("kõnejupp" in Estonian, "Wortgruppierung" in German) in the written transcripts appearing on the screen. Technically, they needed to click on the words that they perceived as occurring at some sort of a break. In essence, the task was to make a binary choice to either place a boundary or not at each consecutive pair of words in an excerpt. No additional instructions on what exactly this break might be were provided. The Estonian listeners were allowed to listen to the excerpts two times, the German listeners were able to listen to the excerpts as many times as they needed.

As this task requires listening to speech excerpts and simultaneously reading written transcripts, it is recognizably difficult for a non-native listener to perform. However, it has already been successfully administered with languages that are typologically far apart in a study by Himmelmann et al. (2018), in which German listeners were asked to chunk speech excerpts from Indonesian languages, and speakers of Indonesian languages were asked to chunk speech excerpts in German. Riesberg et al. (2020) followed a similar procedure with German and Papuan Malay speakers. Both studies yielded interpretable and plausible results. The researchers' justification for this procedure was based on the shared orthographic conventions of the languages.

Estonian orthography is phonemic, and therefore, it should be easily accessible to a German listener/reader. Except for some contrasts in phoneme length, each symbol is encoded by exactly one sound, and most of the graphemes correspond to symbols in German. The survey conducted after the completion of the task indicated that the participants were happy to take part in the study: the average satisfaction on a scale of 0 to 100 was 78.7 (SD = 20.7). 13.3 percent of participants claimed to have difficulties with mapping speech sounds to written words, and 11.1 percent of participants even reported having fun listening to a language that they did not know.

We did not manage to present the lists to equal numbers of participants, as the LMEDS software does not have the option to define different lists of experimental stimuli. Unfortunately, our own solution for extending the LMEDS with this feature did not work properly. Thus, the number of listeners per excerpt varies across the lists, ranging from a total of 6 to a total of 12 listeners per list.

The participants' responses were encoded at the final boundary of every word, using 0 when no boundary was placed and 1 when a boundary was placed. Altogether, the Estonian results consisted of 55,541 data points, and the German results consisted of 47,257 data points (number of words multiplied by the number of listeners). We did not instruct the participants to listen for breaks in the very last words of excerpts, and therefore, the final words of each excerpt were excluded from the evaluation of effects, leaving us with 50,889 data points for the Estonian data and 43,291 data points for the German data.

2.4 Test variables

Four test variables capturing the variation in duration (syllable duration, pause duration), intensity (intensity difference) and F0 (F0 proportion) were automatically extracted from all words in the excerpts. The absolute duration of the last syllable of every word (syllable duration in milliseconds) was taken to index preboundary lengthening. An utterance was defined to be a stretch of fluent speech between silent pauses of 400 ms or longer. Thus, the selected utterances did not contain pauses that were longer than 400 ms. However, they did contain silent and filled pauses shorter than 400 ms (352 instances (0.07%) in a corpus of 4372 words). The duration of these silent and filled pauses was collected as the second durational variable after syllable duration (pause duration in milliseconds).

For the third variable, intensity difference, the intensity as root mean square (RMS) amplitude of the very first and the very last syllable of a word was automatically extracted, and the intensity curve within a word was approximated by subtracting the RMS value of the last syllable from the RMS value of the first syllable (intensity difference). The intensity difference was calculated to index the intensity drop. The larger the intensity difference, the likelier it is that a word contains the intensity drop. A small or negative difference is an indication that a word does not contain an intensity drop.

F0 contours (Hz) were extracted from the excerpts in two passes with the help of the auto-correlation method available in Praat (Boersma & Weenink 2019). During the first pass, F0 tracks were extracted with Praat default settings for the lowest and highest F0, the “floor” and “ceiling” (75 Hz and 600 Hz, respectively). Then, the first and third quartiles of F0 (Q1 and Q3) were calculated for each speaker and recorded in a table. In the second pass, F0 contours were extracted with speaker-specific settings (0.75*Q1 for the floor and 1.5*Q3 for the ceiling). Finally, the resulting F0 contours were smoothed by 4 Hz and quadratically interpolated using the corresponding functions in Praat. Based on the F0 contours, F0 maxima (in Hz) were automatically identified in the vowels of the word-initial lexically stressed syllables. This identification procedure is well justified because, in Estonian, the high tone of the falling pitch accent is most frequently aligned with the first syllable (see, e.g., Asu & Nolan 1999). Therefore, relatively high F0 maxima from the word-initial syllables can be taken to index intonational pitch accents.

For the fourth variable – F0 proportion, F0 maxima were divided with the corresponding utterance's mean F0. As such, the F0 proportion was devised to approximate the height of a pitch accent relative to the utterance's mean F0. F0 proportion was calculated to normalize the speaker-specific and item-specific tonal variation in the utterances. Due to the well-known phenomenon of F0 declination, F0 maxima are higher at the beginnings of the corresponding domains (e.g., IP, clause, or a perceptual speech chunk) than at the ends of these domains (Cooper & Sorensen 1981, Liberman & Pierrehumbert 1984, Yuan & Liberman 2014). Therefore, F0 maxima decrease across the domain also relative to the utterance's mean. In other words, F0 proportion is smaller at the ends of corresponding domains than at the beginnings of these domains. Followingly, the F0 proportion should be smaller at the end of the perceived boundaries than at the non-boundaries if the non-expert perception of a break, or more generally, the perception of a speech chunk relies on the tonal coherence.

The material was also scored for the boundaries of clauses. This scoring was not devised in a particular syntactic framework but followed the functional approach provided in Erelt & Metslang (2017). A clause was defined as consisting of a finite verb together with elements that cluster around the verb and are not finite verbs themselves. Clauses were allowed to also consist of non-constituents, such as disclosures and interjections. In practice, conjunctions served as a frequent cue for the separation of utterances into smaller units of clauses (see rows 7 and 12 in Table 1). For clausal structure, the last word in a clause was encoded as being at the clause boundary.

2.5 Analysis

In our analysis, the continuous variables of syllable and pause duration, intensity difference, and F0 (F0 proportion) function as bottom-up information, whereas clause boundaries function as top-down information. In terms of the impact of continuous signal-based prosodic variables in perceptual chunking, we expected the likelihood of boundary perception to increase

1. together with increasing syllable duration,
2. together with increasing pause duration,
3. together with increasing intensity difference,
4. together with decreasing F0 proportion.

Table 1: Sample of the scoring of clause boundaries in conversational utterances.

Row	Transcription	Translation	Function	Clause boundary
1	*ja*	and	conjunction	no
2	*siis*	then	adverbial	no
3	*käisi-me*	went-we	verb	no
4	*seal*	there	adverbial	no
5	*iisraeli*	Israelian	adverbial	no
6	*muuseum-is*	museum-in	adverbial	yes
7	*kus*	where	conjunction	no
8	*see*	this	subject	no
9	*suur*	big	subject	no
10	*makett*	maquette	subject	no
11	*oli*	was	verb	yes
12	*mis*	which	conjunction	no
13	*oli*	was	verb	no
14	*päris*	pretty	predicative	no
15	*võimas*	awesome	predicative	yes

We predicted that the perception of both types of information would be modulated by the listener's linguistic background (familiar vs. unfamiliar) such that the effects of prosodic variables would be larger for German than for Estonian listeners and that the effect of clause boundaries would be larger for Estonian than for German listeners.

The effects of clause boundaries, syllable duration, intensity difference, F0 proportion and pause duration were estimated in relation to the language background in the general linear mixed effects regression analysis as provided in the `lme4` package (Bates et al. 2015) in R (R Core Team 2018). We defined five predictors of the binomially distributed response variable:

1. an interaction between clause boundaries and language,
2. an interaction between syllable duration and language,
3. an interaction between pause duration and language,
4. an interaction between intensity difference and language,
5. an interaction between F0 proportion and language.

Pause and syllable durations were logarithmically transformed with the base of 10. To maintain the interpretability and comparability of the slopes, all continuous variables were z-scored before entering the regression analysis. The generalized linear mixed effects model was defined to contain the number of listeners as an exposure variable because the four lists of excerpts in the Estonian experiment and the nine lists of excerpts in the German experiment were exposed to different numbers of listeners. The random effects structure included random slopes for listeners because we reasoned that listeners are highly likely to vary in their sensitivity to the clausal structure, syllable duration, pause duration, intensity difference and F0 proportion. We also included random slopes for excerpts because they originated from the conversations of 10 different speakers and displayed considerable and systematic variation in speech rhythm, intensity, and melody. The converging model fit was obtained by using the `optimx` optimizer (Nash 2014, Nash & Varadhan 2011).

3 Results

3.1 The impact of prosodic cues on non-native speech chunking

The aim of the analyses was to determine the impact of phonetic variation of duration, intensity and F0 as bottom-up information in non-native speech chunking. Before proceeding to the statistical evaluation, the explanatory variables were checked for correlations (see Table 2).

Table 2: Correlations between the explanatory variables as estimated by Pearson's r coefficient. The significance stars indicate how likely they are to be found in the whole population, given the sample means. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

	Clause	Syl. dur.	Int. dif.	F0 prop.
Clause				
Syllable dur.	0.08^{***}			
Intensity dif.	0.12^{***}	-0.02		
F0 prop.	0.04^{**}	0.04^{*}	0.19^{***}	
Pause dur.	-0.01	0.07	-0.06	0.03

The correlations between the selected variables in Table 2 are very close to zero. This indicates that they are appropriate as explanatory variables for the multiple regression analysis with mixed effects. The results of the analysis are presented in Table 3. The column "Est." contains the log odd estimates of the fixed effects clause, syllable duration, intensity difference, pause duration and F0 proportion in interaction with language. The third and the fourth column give the 95% confidence intervals of the estimates. The *t*-values and *p*-values can be found in the last two columns. The *p*-values are given together with the significance codes (asterisks).

Table 3: Log odd estimates and significance of the standardized variables in predicting boundary perception. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

	Est.	2.5%	97.5%	t	p
(Intercept)	−12.22	−12.99	−11.46	−31.35	0.00***
Language [Ger]	1.84	0.97	2.71	4.15	0.00***
Clause [yes]	3.18	2.65	3.71	11.72	0.00***
Syllable dur.	0.04	−0.21	0.29	0.3	0.77
Intensity dif.	0.06	−0.31	0.44	0.34	0.73
Pause dur.	0.53	0.27	0.79	4.06	0.00***
F0 prop.	0.27	0	0.54	1.98	0.05*
Language [Ger]:Clause [yes]	−2.57	−3.11	−2.04	−9.41	0.00***
Language [Ger]:Syllable dur.	0.22	0.01	0.42	2.1	0.04*
Language [Ger]:Intensity dif.	0.32	0.08	0.56	2.57	0.01**
Language [Ger]:Pause dur.	0.06	−0.15	0.26	0.54	0.59
Language [Ger]:F0 prop.	0.16	−0.04	0.37	1.55	0.12
AIC					5770.47
R^2 (fixed effects)					0.15
R^2 (all effects)					0.78

The positive values of the log odd estimates indicate an increase in the probability of boundary perception, whereas the negative values suggest a decrease in the probability of boundary perception. Given that the variables were standardized before entering the regression analysis, the estimates enable us to see that the presence of a clause boundary is the factor that has the most profound effect on boundary perception. This is followed by the effect of the interaction between the language and clause and the main effect of the language. The lower-ranking effects stem from the signal-based prosodic variables. The main effect of the language is followed by the main effect of pause duration. The next strongest effect is the intensity difference in the interaction with language. This is followed by the main effect of F0 proportion. Finally, syllable duration also contributes to the boundary perception in the interaction with language. The main effects of syllable duration and intensity difference, and the interactions between language and pause duration and between language and F0 proportion did not turn out significant. The results of the linear-mixed effects regression analysis are illustrated in the effect plots in Figure 1. These plots highlight the predicted influences of clause boundaries, syllable duration, intensity difference, pause duration and F0 proportion on boundary perception.

Figure 1A further demonstrates how the significant main effect of clause boundaries is modulated by the significant interaction between clause boundaries and language background. In particular, we can see that the Estonian listeners are strongly affected by the presence of a clause boundary whereas the German listeners are insensitive to the presence of clause boundaries (compare blue points and whiskers to red points and whiskers). Figure 1B demonstrates that increasing duration of the last syllable contributes to the perception of a boundary for German (see the blue line and confidence intervals that are not overlapping from left to right) but not for Estonian listeners (see the red line and the red confidence intervals that are overlapping from left to right along the probability function). Similarly, Figure 1C indicates that the probability of hearing a boundary increases together with increasing intensity difference for German listeners (see the blue line and confidence intervals that are not overlapping from left to right) but not for Estonian listeners (see the red line and the red confidence intervals that are overlapping from left to right along the probability function). Figures 1D and 1E underscore the main effects of pause duration and F0 proportion. We can readily observe that regardless of the listener's language background, the probability of boundary perception increases as the pause duration and F0 proportion increase (see the rising probability functions and non-overlapping confidence intervals in blue and red from left to right along the probability functions).

3.2 Interrater agreement

To establish the interrater agreement, we calculated Fleiss' κ scores between the Estonian and German listeners according to the lists of excerpts (see Table 4).

The κ scores in Table 4 show fair agreement within Estonian listeners and within German listeners. While Estonian listeners of Lists 1 and 2 perform moderately, the scores for other lists remain below 40, yielding an average κ score of 0.38 for Estonians. The average κ score for German listeners is 0.28, also indicating fair agreement. It was not possible for us to calculate the κ scores between the Estonian and German listeners because the excerpts were distributed among the different lists (among four lists for Estonians and nine lists for Germans).

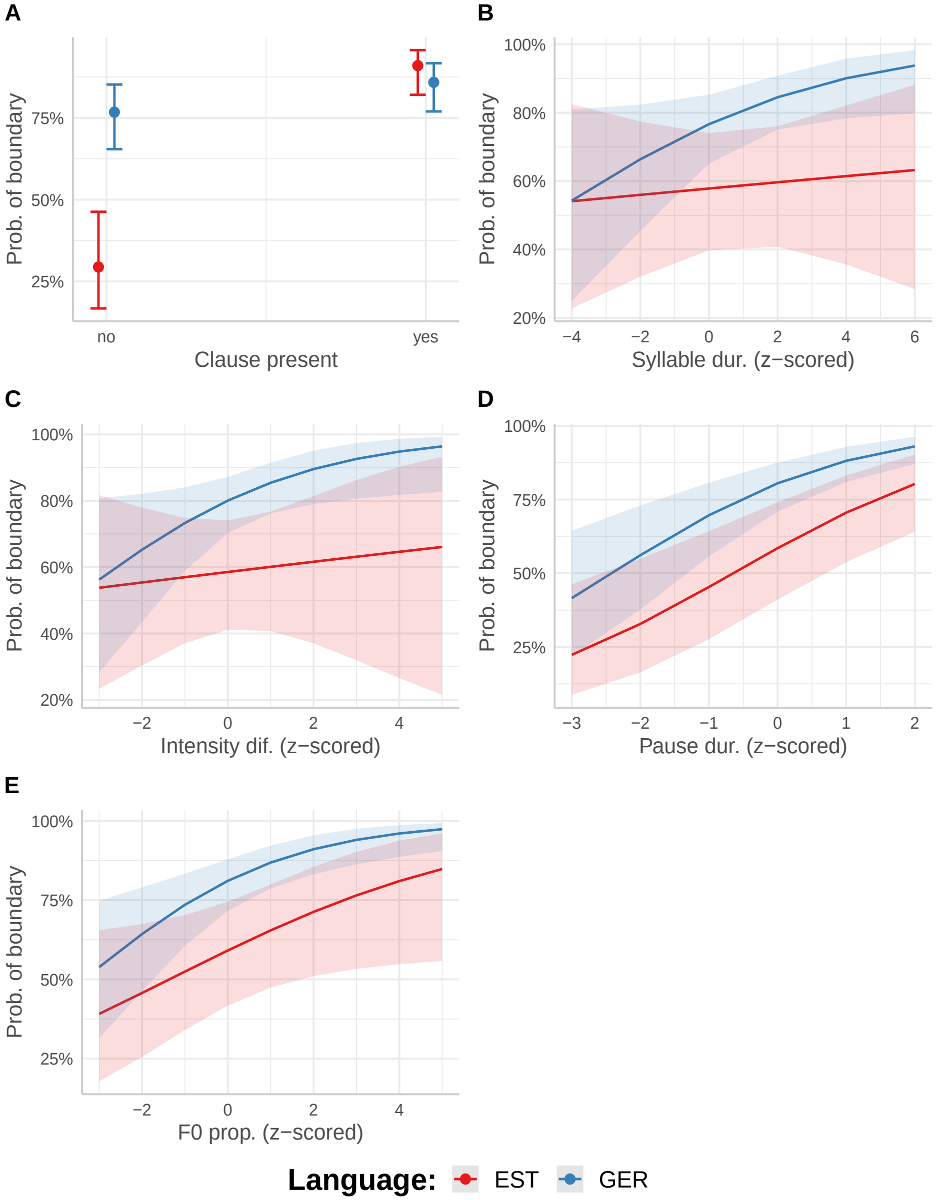


Figure 1: Predicted probabilities of boundary perception as a function of clause boundaries (A), syllable duration (B), intensity difference (C), pause duration (D) and F0 proportion (E) while holding other variables constant. The shadowed bands around the lines represent 95% confidence intervals of the estimates. The change in the probability function is significant when the confidence intervals do not overlap from left to right along the probability function.

Table 4: Fleiss' κ scores for boundaries in the familiar (Estonian) and unfamiliar (German) language conditions. N: Number of listeners. The κ values between 0–0.20 indicate *slight* agreement, 0.21–0.40 suggest *fair* agreement, 0.41–0.60 indicate *moderate* agreement, 0.61–0.80 indicate *substantial* agreement, and 0.81–1 suggest *almost perfect* agreement (see Landis & Koch 1977).

List	N	κ	95% CI	z	Agreement
Estonian					
1	13	0.47	(0.46, 0.47)	142.32	moderate
2	9	0.41	(0.40, 0.42)	85.67	moderate
3	14	0.39	(0.38, 0.39)	126.14	fair
4	11	0.27	(0.26, 0.28)	68.37	fair
Mean (SD):			0.38 (0.08)		(fair)
German					
1	10	0.29	(0.28, 0.30)	45.05	fair
2	10	0.34	(0.33, 0.36)	51.91	fair
3	9	0.33	(0.31, 0.34)	44.12	fair
4	11	0.25	(0.24, 0.26)	42.17	fair
5	13	0.25	(0.24, 0.26)	42.17	fair
6	12	0.28	(0.27, 0.29)	56.33	fair
7	7	0.23	(0.22, 0.24)	42.42	fair
8	6	0.32	(0.30, 0.34)	28.17	fair
9	12	0.27	(0.26, 0.28)	51.07	fair
Mean (SD):			0.28 (0.04)		(fair)

Therefore, we decided to investigate the perceptual chunks to see whether they show any similarities between the native and non-native speakers. The results of the regression analysis have strongly indicated that for the Estonian listeners, the boundaries of chunks correspond with clause boundaries. Additionally, they are guided by pause duration and F0 proportion. The German listeners, in contrast, are not affected by clause boundaries and rely more strongly on the acoustic characteristics of words (syllable duration, intensity difference, pause duration and F0 proportion). Therefore, we decided to investigate some lexical and prosodic characteristics of the chunks that were identified by the German and Estonian listeners. Firstly, we examined the length of the chunks in terms of duration (in milliseconds) and the number of words. There is an idea that chunking processes could be constrained by the capacity of working memory, which has been frequently measured in how many words a person is able to recall (Green 2017). The finding is that the working memory mostly spans from five to seven words (sometimes even nine words, Miller 1956). We speculated that the German listeners might be stronger constrained by the memory capacity than the Estonian listeners because language processing and memory of the Estonian listeners are supported by the semantic and syntactic information that is inaccessible to the German listeners. So, we expected the duration of non-clausal chunks that were perceived by German listeners to conform stronger with the memory constraint than the duration of the clausal chunks that were identified by Estonian listeners. In particular, we expected non-clausal chunks to be shorter and less variable than the clausal chunks.

Secondly, we analysed the lexical constituency of the chunks. Concerning words, we expected that the chunks identified by the Estonian listeners are more likely to begin with conjunctions and the so-called clausal connectors (e.g. *et*, ‘that’; *aga*, ‘but’; *kui*, ‘if/when’; etc.) than the chunks identified by the Germans. This is because conjunctions signal the beginning of a new clause (in our analysis) and only the native speakers have access to this syntactic information. Thus, it is not likely that the German listeners would consistently identify conjunction-initial chunks. Finally, we explored the tonal coherence of the perceptual chunks. As discussed in the Introduction, the tonal coherence can be approximated by the decline of F0 across the respective domain (e.g., IP, clause or perceptual chunk). Thus, we visually estimated the degree of tonal coherence of the perceptual chunks by observing the averaged F0 contours of the native and non-native language chunks. We speculated that the non-native chunks (non-clausal chunks) exhibit tonal coherence to a larger degree than the native chunks (clausal chunks) because the German listeners were stronger guided by the bottom-up prosodic cues than were the Estonian listeners.

3.3 Any shared characteristics between the native and non-native chunks?

We examined the chunks identified by the Estonian and German listeners considering the chunks’ length (in duration and number of words, Table 5), lexical characteristics (Table 6) and tonal coherence (Figure 2).

The averages of duration and length in words in Table 5 indicate that the perceptual chunks do not differ in duration or the number of words between the two language groups. In other words, listeners with Estonian and German backgrounds identify chunks of the same length and size. The difference is that the chunks identified by Estonian listeners are more likely to form a clause than the chunks identified by the German listeners.

Table 5: Lengths of chunks in German and Estonian listeners as estimated by duration (ms) and number of words in chunks.

	Av. duration (ms)		Length in words	
Language group	Mean	SD	Mean	SD
Estonian	1452	850	5.85	3.35
German	1417	749	5.86	3.24

For the lexical characteristics in Table 6, we identified words that appeared most frequently in the first, second, third and final positions in the chunks. The aim was to see if the lexical content of the chunks differs between the two language groups.

Table 6 reveals no differences in the lexical constituency of Estonian and German chunks. The word frequencies reflect the nature of spontaneously spoken Estonian, in which the connectors (*et* ‘that’, *ja* ‘and’) and the pronouns (*ma* ‘I’, *see* ‘this’) have the highest frequency (see Lippus 2019).

Furthermore, we investigated the tonal coherence of the perceptual chunks that were identified by German and Estonian listeners. For this, we extracted F0 contours of each excerpt identified by each listener and categorized them based on their position within the excerpt: (i) at the beginning of the excerpts, that is, first chunk, (ii) following the first chunk, that is second chunk, (iii) at the end of the chunk, that is final, (iv) and all others between the second and the last chunk within the excerpt. There were 3999 three-chunk excerpts (46.9 percent of all the chunked excerpts), 2615 four-chunk excerpts (30.6 percent of all the chunkings) and only 1099 two-chunk excerpts (12.9 percent of all the chunkings). F0 contours of the perceptual chunks were then time-normalized by extracting 32 F0 measures, equally distributed within a respective perceptually identified chunk. The 32 measurements of F0 were then averaged by their position (see Figure 2). The different panels in Figure 2 enable us to follow the decline of F0 in the excerpt-initial chunks, in the chunks of second position, the chunks of excerpt-medial position, and the chunks of the excerpt-final position.

We can observe a continuous decline in F0 over the entire excerpt but also over the chunks identified at the different positions in the excerpts. Tonally, the chunks identified by German and Estonian listeners are comparable, and no major differences occur.

Table 6: The 10 most frequent words in the first, second, third and final positions of chunks identified by Estonian and German listeners. FR: frequency ranking

FR	First pos.		Second pos.		Third pos.		Last pos.	
Estonian								
1	*et*	‘that’	*siis*	‘then’	*on*	‘is’	*et*	‘that’
2	*ja*	‘and’	*ei*	‘no’	*ei*	‘no’	*noh*	‘well, uhm’
3	*siis*	‘then’	*ma*	‘I’	*oli*	‘was’	*on*	‘is’
4	*ma*	‘I’	*see*	‘this’	*et*	‘that’	*siis*	‘then’
5	*aga*	‘but’	*on*	‘is’	*me*	‘we’	*see*	‘this’
6	*see*	‘this’	*oli*	‘was’	*see*	‘this’	*ka*	‘too’
7	*kui*	‘if, when’	*et*	‘that’	*nagu*	‘like’	*oli*	‘was’
8	*või*	‘or’	*ta*	‘(s)he’	*ma*	‘I’	*jah*	‘yes’
9	*ei*	‘no’	*me*	‘we’	*seal*	‘there’	*ja*	‘and’
10	*mingi*	‘some’	*seal*	‘there’	*kui*	‘if, when’	*seda*	‘this [PART]’
German								
1	*et*	‘that’	*siis*	‘then’	*on*	‘is’	*et*	‘that’
2	*ja*	‘and’	*ma*	‘I’	*ei*	‘no’	*see*	‘this’
3	*siis*	‘then’	*ei*	‘no’	*oli*	‘was’	*ja*	‘and’
4	*ei*	‘no’	*et*	‘that’	*et*	‘that’	*siis*	‘then’
5	*on*	‘is’	*see*	‘this’	*see*	‘this’	*nagu*	‘like’
6	*ma*	‘I’	*on*	‘is’	*ma*	‘I’	*mingi*	‘some’
7	*see*	‘this’	*oli*	‘was’	*me*	‘we’	*on*	‘is’
8	*oli*	‘was’	*seal*	‘there’	*kui*	‘if, when’	*seda*	‘this [PART]’
9	*noh*	‘well, uhm’	*ja*	‘and’	*nagu*	‘like’	*oli*	‘was’
10	*aga*	‘but’	*kui*	‘if, when’	*siis*	‘then’	*noh*	‘well, uhm’

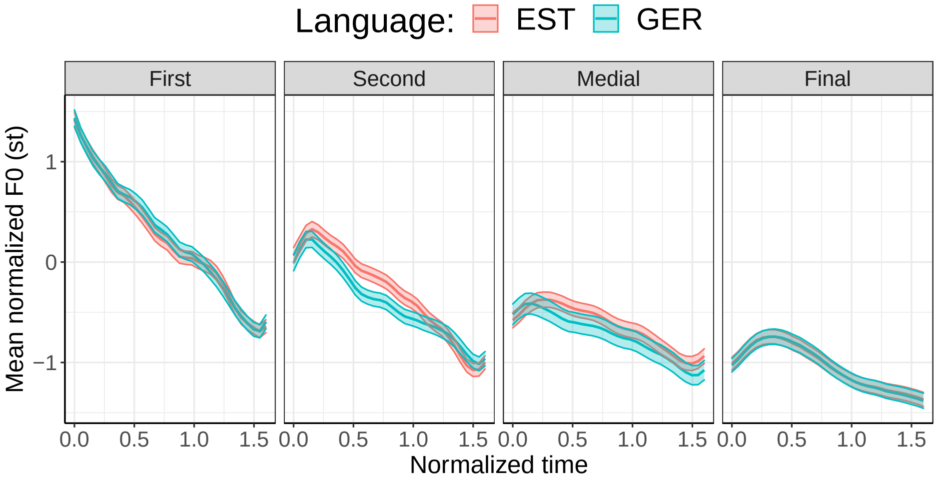


Figure 2: Time-normalized F0 contours converted into semitones (st) by position in excerpts and by language background of listeners. “First” refers to the excerpt-initial position and “Second” to the second position in an excerpt. “Medial” incorporates all other positions except the final position, and “Final” indicates the excerpt-final chunks.

4 Discussion

This study set out to investigate the impact of signal-driven prosodic cues on speech boundary perception in listeners of an unfamiliar language. We employed an RPT experiment in which German listeners unfamiliar with the Estonian language were asked to chunk spontaneous utterances spoken in Estonian. The results of the experiment were compared to the results of a previous experiment in which Estonian listeners were asked to perform a similar task listening to the same speech excerpts (Ots & Taremaa 2022). We examined the duration of word-final syllables, the duration of pauses, intensity curves and F0 (F0 maxima relative to average F0 of respective sentences) as set against the clausal structure at the chunk boundaries identified by German and Estonian listeners.

The results show that German listeners appear to use all the phonetic cues of syllable duration, intensity, pause duration and F0, and to ignore the clausal information. In contrast, Estonian listeners mostly utilize top-down information, as they largely relied on clause boundaries in the chunking task. After the clausal information, Estonian listeners also used the phonetic cues of pause duration and F0 but not syllable duration and intensity difference.

More specifically, the results demonstrate for German listeners that the probability of boundary perception increased together with increasing duration of the word-final syllable. The longer syllable duration corresponding with the chunk boundary resembles the well-known prosodic boundary cue – the pre-boundary or phrase-final lengthening (Berkovits 1994, Fon et al. 2011, Nakai et al. 2009, Petrone et al. 2017, Wightman et al. 1992; for German, see also Schubö & Zerbian 2023 [this volume], Huttenlauch et al. 2023 [this volume], and Wellmann et al. 2023 [this volume]). Thus, it seems that the German listeners are guided by phrase-final lengthening while chunking an unfamiliar language. Furthermore, the analysis indicates that boundary perception became likelier among the German listeners as the intensity difference between the first and last syllable in the word increased. This suggests that German listeners interpreted the intensity drop as an additional cue for a chunk boundary.

Although pauses are usually infrequent in conversational utterances (Biron et al. 2021), they are known to be accessible and reliable cues for boundary perception (Himmelmann et al. 2018, Riesberg et al. 2020, Petrone et al. 2017). In our study, we observe that listeners of both familiar and unfamiliar language backgrounds benefited from the presence of longer, rather than shorter, pauses: the longer the pause, the likelier the perception of a chunk boundary. Similarly, both language groups benefited from variation of F0. However, the tonal cue was interpreted in the opposite direction from what was predicted. The higher the word-initial stressed syllable was relative to the sentence's mean F0, the likelier it was that the listener would perceive a boundary after this word.

At first sight, the result concerning intonation is somewhat puzzling. As an explanation, we consider that the F0 maxima in our materials index a sort of rising boundary tone and not pitch accents. Theoretically, the stressed syllable preceding a rise is pitched low, and the F0 maxima in the final unstressed syllables should index the rise right before the phrase boundary. The lexical makeup of the identified chunks in Table 6 demonstrates that the most frequent words at the ends of chunks were monosyllabic words. Monosyllabic words are considered to carry lexical stress, but they tend to become reduced and unstressed in unaccented positions of spoken utterances (Lehiste 1960: 54). As such, they may well serve as carriers of the phrase-final tonal rise. Thus, our listeners, irrespective of their language background, interpreted increasing F0 contour as a cue for a chunk boundary. As such, the result corroborates the findings in Petrone et al. (2017) and in Kentner & Féry (2013), who have found for German that the F0 in the first and last syllables of phrase-final words at the IP-medial positions is high, that is, the IP-medial phrase-final words have a strong tendency to carry a tonal rise. Our study, together with Petrone et al. (2017), establishes that tonal rises are interpreted as boundary cues also in the perception of spontaneous speech natively and non-natively.

In comparing the two language groups, we discovered that while phrase-final lengthening and intensity drop functioned as boundary cues for German listeners, they did not for Estonian listeners. It is possible that this difference might relate to the differing prosodic profiles of these languages. For example, steep F0 falls accompanied by a deep intensity drop are quite common for German declarative sentences (Peters 1999, Ulbrich 2002). Thus, German listeners might be attuned to hearing large intensity drops accompanied by tonal falls as boundary cues. Similarly, phrase-final lengthening is most frequently attested in German and English. However, the lengthened segments signal the three-way quantity contrast of phonological feet that distinguishes between morpho-lexical functions in Estonian (Eek 1990, Lehiste 1960, 1997). Although the phonological variation of duration does not directly confine the phenomenon of phrase-final lengthening in production, Estonian listeners might nevertheless concentrate on aspects of segmental lengthening differently from German listeners. Thus, the results on intensity drop and pre-boundary lengthening indicate that the crosslinguistic applicability of prosodic boundary cues depends on the prosodic characteristics of the crossed languages.

Clause boundaries, phrase-final lengthening, intensity drop and rising boundary tone performed well in explaining the distribution of boundary marks in the logistic mixed-effects analysis, but the concordance within the two groups of participants showed that the listeners demonstrated only fair agreement in identifying the presence of a boundary. The Fleiss' κ scores compared to the κ scores reported in previous studies were considerably lower (see, e.g., Himmelmann et al. 2018, Riesberg et al. 2020). This holds true especially for the Estonian listeners who attended to their native language. On the one hand, this result might arise from the nature of the materials the participants were asked to listen to. The utterances were extracted from a corpus of dialogues that were held among friends or acquaintances on a freely chosen topic. Although they were recorded in an unnatural recording situation (in a professional sound-attenuated recording studio), these utterances represent highly conversational speech. The low agreement numbers most likely reflect the high acoustic variability characteristic of conversational speech. Also, the selected utterances probably display several different combinations of acoustic boundary cues in which pauses, pause duration, pre-boundary lengthening, intensity drop, and increasing F0 contour are produced at varying strengths. Rising F0 movement is usually accompanied by a decrease in intensity difference. As such, the rising boundary cue might counteract the cue of intensity drop. On the other hand, the low concordances suggest that listeners vary greatly in their cue weighting. For example, Baumann & Winter (2018) found that German listeners in a similar chunking task were divided into two groups: those who attend to pitch-related cues (such as pitch accent type, mean and maximum F0) and those who instead rely on duration and lexical and syntactic information. Most likely, the participants of the experiment made sense of numerous combinations of boundary cues in many different ways, which also explains the low agreement scores.

In the final part of the analysis, we compared the lexical and acoustic characteristics of the speech chunks identified by the German and Estonian listeners. The native and non-native speech chunks displayed a number of shared characteristics. Specifically, the chunks were comparably long in duration and in the number of words. They also displayed very similar lexical variation, common for spontaneous speech in general. More importantly, the average F0 contours demonstrate that the speech chunks identified by both language groups conform to the concept of tonal coherence. Regardless of position in the excerpts, F0 was gradually declining across the native as well as non-native speech chunks. Thus, the chunks identified by the German and Estonian listeners differed from each other neither prosodically nor lexically.

The Estonian chunks, however, corresponded more frequently with the syntactic clauses. To stay within the boundaries of the current study, we must refrain from further examination of the chunks that the German listeners identified. However, we find it very interesting that the German participants clearly found types of speech chunks that are not clauses but show prosodic coherence and high comparability with the clauses detected by native listeners. For future research, we propose to investigate what types of chunks German listeners identify in terms of semantic and pragmatic coherence and whether these could be helpful for language learners when decoding a second language.

Overall, the study provides evidence that the two language groups – German and Estonian listeners – employed longer pauses and rising F0 contour in a speech chunking task. In other words, we have found crosslinguistic application of pausing and F0. As non-native listeners, Germans additionally utilized preboundary lengthening and intensity drop. Thus, while German listeners made use of all acoustic variables we investigated here, Estonian listeners applied only a few of them and relied mainly on the presence of clause boundaries.

We categorized phonetic variables (duration, intensity and F0) as bottom-up information and clause boundaries as top-down information. We predicted less influence from clausal information but more influence from signal-based prosodic information for German listeners than for Estonian listeners. As discussed above, the results support this prediction. As expected, in chunking Estonian speech, German listeners unfamiliar with the Estonian language make use of bottom-up information only, whereas Estonian listeners mostly utilize top-down information, as they largely relied on clause boundaries in the chunking task. This outcome runs counter to the results in Riesberg et al. (2020) and demonstrates that the production of prosody in Estonian spontaneous speech is not too tightly bound to the clausal structure. Nevertheless, the results reflect well on the bottom-up and top-down processing mechanisms.

Clearly, when a listener has no knowledge of a language, prosodic boundary cues are the primary source of information for making sense of speech in an unfamiliar language. Native listeners, however, mainly employ semantic and syntactic knowledge, that is, top-down information, but, as we have seen, benefit from prosodic information as well. We speculate that the role of prosodic information is even greater but in the type of RPT task, it is flooded with semantic and syntactic information which emerges from lexical sources. Therefore, the role of prosody in the earliest stages of spoken language processing might be better established by using more sensitive methods being able to tap into the ongoing decoding processes (see Wellmann et al. 2023 [this volume] for boundary perception in infants). Nevertheless, our study of non-native listeners in comparison to the previous study of native listeners has successfully demonstrated both bottom-up and top-down effects in the processing of spontaneous speech.

We probably see top-down processing somewhat overriding bottom-up processing in native speech processing. This is understandable because top-down processing, together with prediction, is an efficient way to reduce the cognitive load, as it enables one to avoid processing every single aspect of information available in the environment (Bar et al. 2006, Clark 2016, Engel et al. 2001). We believe that the phenomenon of top-down processing also explains the results of previous phonetic perception experiments in which boundary perception in native listeners has been shown to be mediated mainly by syntactic and lexical variables (e.g., Cole et al. 2010, Christodoulides et al. 2018, Baumann & Winter 2018). To demonstrate the impact and functions of bottom-up information – signal-driven prosodic boundary cues in particular – for native listeners, future studies should involve more rigorous research techniques that can assess on-going comprehension.

5 Conclusion

In this study, we investigated the impact of signal-driven prosodic cues on chunking excerpts of a natural language. For this, we utilized RPT methodology and asked non-native listeners (Germans) to identify speech chunks in excerpts spoken in an unfamiliar language (Estonian). We examined the acoustic variation at the boundaries of chunks identified by German listeners with reference to chunk boundaries detected in the same excerpts by Estonian listeners in an earlier experiment. The results show that German listeners, having no access to the semantic-syntactic structure of Estonian, largely rely on signal-driven prosodic information and utilize syllable duration, intensity curves, pause duration and rising F0 contour when dividing a continuous stream of speech into smaller chunks. Estonians, on the contrary, rely mainly on the presence of clause boundaries, but they additionally apply pause duration and rising F0 contour for the identification of speech chunks. The results demonstrate the importance of signal-driven prosodic boundary cues in bottom-up processing of spoken language and highlight the interaction between bottom-up processing (sensory input from speech acoustics) and top-down processing (linguistic knowledge about clause structure) in native speech comprehension.

Acknowledgments

We are extremely thankful to the Estonian volunteers and the German participants who took part in our experiments. We further appreciate the warm and supportive audience of the workshop “Prosodic boundary phenomena” at the 43rd annual meeting of the DGfS (Deutsche Gesellschaft für Sprachwissenschaft), who strongly motivated a study with non-native listeners.

Funding information

This work was supported by research funding awarded to the first author by Fritz Thyssen Stiftung in Germany (10.18.2.040SL, “Planning sentences and sentence intonation cross-linguistically”) and by funding from the European Union through the European Regional Development Fund (Centre of Excellence in Estonian Studies) and from the research fund of Kadri, Nikolai, and Gerda Rõuk, both of which were awarded to the second author.

References

Asu, Eva Liina & Francis Nolan. 1999. The effect of intonation on pitch cues to the Estonian quantity contrast. In *Proceedings of the 14th ICPhS*, 1873–1876. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1873.pdf.

Bar, Moshe, Karim S. Kassam, Avniel Singh Ingh Ghuman, Jasmine Boshyan, Ann Marie Schmidt, Anders M. Dale, Matti S. Hämäläinen, Ksenija Marinkovic, Daniel L. L. Schacter, Bruce Robert Rosen & Eric Halgren. 2006. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences USA* 103(2). 449–454. DOI: 10.1073/pnas.0507062103.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: 10.18637/jss.v067.i01.

Baumann, Stefan & Bodo Winter. 2018. What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics* 70. 20–38. DOI: 10.1016/j.wocn.2018.05.004.

Berkovits, Rochele. 1994. Durational effects in final lengthening, gapping and contrastive stress. *Language and Speech* 37. 237–250. DOI: 10.1177/002383099403700302.

Biron, Tirza, Daniel Baum, Dominik Freche, Nadav Matalon, Netanel Ehrmann, Eyal Weinreb, David Biron & Elisha Moses. 2021. Automatic detection of prosodic boundaries in spontaneous speech. *PLOS ONE* 16(5). 1–21. DOI: 10.1371/journal.pone.0250969.

Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer.* http://www.praat.org/.

Bois, John W. Du, Susanna Cumming, Stephan Schuetze-Coburn & Danae Paolino (eds.). 1992. *Discourse transcription*, vol. 4. Santa Barbara: Department of Linguistics, University of California.

Breen, Mara, Laura C. Dilley, John Kraemer & Edward Gibson. 2012. Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory* 8(2). 277–312. DOI: 10.1515/cllt-2012-0011.

Buhmann, Jeska, Johanneke Caspers, Vincent J. van Heuven, Heleen Hoekstra, Jean-Pierre Martens & Marc Swerts. 2002. Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the spoken Dutch corpus. In *Proceedings of the third international conference on Language Resources and Evaluation (LREC'02).* http://www.lrec-conf.org/proceedings/lrec2002/pdf/96.pdf.

Chafe, Wallace. 1987. Cognitive constraints on information flow. In Russell S. Tomlin (ed.), *Coherence and Grounding in Discourse: Outcome of a Symposium* (Typological Studies in Language 11), 21–51. Amsterdam: Benjamins. DOI: 10.1075/tsl.11.03cha.

Christiansen, Morten H. & Nick Chater. 2016. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences* 39. e62. DOI: 10.1017/S0140525X1500031X.

Christodoulides, George, Anne Catherine Simon & Ivana Didirkova. 2018. Perception of prosodic boundaries by naive and expert listeners in French. Modelling and automatic annotation. In *Proceedings of Speech Prosody*, 641–645. DOI: 10.21437/SpeechProsody.2018-130.

Clark, Andy. 2016. *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780190217013.001.0001.

Cole, Jennifer, Yoonsook Mo & Soondo Baek. 2010. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes* 25(7–9). 1141–1177. DOI: 10.1080/01690960903525507.

Cole, Jennifer, Yoonsook Mo & Mark Hasegawa-Johnson. 2011. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology* 1(2). 425–452. DOI: 10.1515/labphon.2010.022.

Cooper, William E. & John M. Sorensen. 1981. *Fundamental frequency in sentence production.* New York: Springer.

Couper-Kuhlen, Elizabeth. 2001. Intonation and discourse: Current views from within. In Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton (eds.), *The handbook of discourse analysis*, 13–34. Malden, MA: Blackwell. DOI: 10.1002/9780470753460.ch2.

Cutler, Anne, Delphine Dahan & Wilma van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40(2). 141–201. DOI: 10.1177/002383099704000203.

Dahan, Delphine & Fernanda Ferreira. 2019. Language comprehension: Insights from research on spoken language. In Peter Hagoort (ed.), *Human language: From genes and brains to behavior*, 21–33. Cambridge, MA: MIT Press.

de la Cruz-Pavía, Irene, Janet F. Werker, Eric Vatikiotis-Bateson & Judit Gervain. 2019. Finding phrases: The interplay of word frequency, phrasal prosody and co-speech visual information in chunking speech by monolingual and bilingual adults. *Language and Speech* 63(2). 264–291. DOI: 10.1177/0023830919842353.

Duez, Danielle. 1985. Perception of silent pauses in continuous speech. *Language and Speech* 28(4). 377–388. DOI: 10.1177/002383098502800403.

Eek, Arvo. 1990. Units of temporal organisation and word accents in Estonian. *Linguistica Uralica* 26(4). 251–263.

Engel, Andreas K., Pascal Fries & Wolf Singer. 2001. Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience* 2(10). 704–716. DOI: 10.1038/35094565.

Erelt, Mati & Helle Metslang. 2017. *Eesti keele süntaks [eng. Estonian syntax]* (Eesti keele varamu). Tartu: Tartu Ülikooli Kirjastus.

Féry, Caroline & Shinichiro Ishihara. 2009. How focus and givenness shape prosody. In Malte Zimmermann & Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 36–63. Oxford: Oxford University Press.

Fon, Janice, Keith Johnson & Sally Chen. 2011. Durational patterning at syntactic and discourse boundaries in Mandarin spontaneous speech. *Language and Speech* 54(Pt 1). 5–32. DOI: 10.1177/0023830910372492.

Fujisaki, Hiroya. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (ed.), *The productions of speech*, 39–55. Heidelberg: Springer.

Fujisaki, Hiroya & Keikichi Hirose. 1982. Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. In *Preprint of papers. Working group on intonation, 13th International Congress of Linguistics*, 57–70.

Green, Clarence. 2017. Usage-based linguistics and the magic number four. *Cognitive Linguistics* 28(2). 209–237. DOI: 10.1515/cog-2015-0112.

Halliday, Michael A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244. DOI: 10.1017/S0022226700016613.

Himmelmann, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology* 35(2). 207–245. DOI: 10.1017/S0952675718000039.

Huttenlauch, Clara, Marie Hansen, Carola de Beer, Sandra Hanne & Isabell Wartenburger. 2023. Age effects on linguistic prosody in coordinates produced to varying interlocutors: Comparison of younger and older speakers. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 157–192. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777534.

Keating, Patricia A. & Stefanie Shattuck-Hufnagel. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 101. 112–156. https://escholarship.org/uc/item/1qf5f44k.

Kentner, Gerrit & Caroline Féry. 2013. A new approach to prosodic grouping. *The Linguistic Review* 30. 277–311. DOI: 10.1515/tlr-2013-0009.

Ladd, D. Robert. 1988. Declination "reset" and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America* 84(2). 530–544. DOI: 10.1121/1.396830.

Ladd, D. Robert. 2008. *Intonational phonology*. Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511808814.

Landis, J. Richard & Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1). 159–174. DOI: 10.2307/2529310.

Langus, Alan, Erika Marchetto, Ricardo Augusto Hoffmann Bion & Marina Nespor. 2012. Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language* 66(1). 285–306. DOI: 10.1016/j.jml.2011.09.004.

Lehiste, Ilse. 1960. Segmental and syllabic quantity in Estonian. *American Studies in Uralic Linguistics*. Uralic and Altaic series (1). 21–82.

Lehiste, Ilse. 1997. Search for phonetic correlates in Estonian prosody. In Ilse Lehiste & Jaan Ross (eds.), *Estonian prosody: Papers from a symposium*, 11–35. Tallinn: Institute of Estonian Language.

Liberman, Mark Y. & Janet B. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In Mark Aronoff, Richard T. Oehrle, Frances Kelley & Bonnie W. Stephens (eds.), *Language and sound structure: Studies in phonology presented to Morris Halle by his teacher and students*, 157–233. Cambridge, MA: MIT Press.

Lippus, Pärtel. 2019. *Foneetikakorpuse sagedussõnastik*. DOI: 10.15155/re-62.

Lippus, Pärtel, Tuuli Tuisk, Nele Salveste & Pire Teras. 2016. *Phonetic corpus of Estonian spontaneous speech v.1.0.0*. DOI: 10.15155/1-00-0000-0000-0000-00074L.

Mahrt, Tim. 2016. *LMEDS: Language markup and experimental design software*. Computer Program.

Mattys, Sven L., Laurence White & James F. Melhorn. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology. General* 134(4). 477–500. DOI: 10.1037/0096-3445.134.4.477.

Miller, George A. 1956. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2). 81–97. DOI: 10.1037/h0043158.

Nakai, Satsuki, Sari Kunnari, Alice E. Turk, Kari Suomi & Riikka Ylitalo. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37. 29–45. DOI: 10.1016/j.wocn.2008.08.002.

Nash, John C. 2014. On best practice optimization methods in R. *Journal of Statistical Software* 60(2). 1–14. DOI: 10.18637/JSS.V060.I02.

Nash, John C. & Ravi Varadhan. 2011. Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software* 43(9). 1–14. DOI: 10.18637/JSS.V043.I09.

Ordin, Mikhail, Leona Polyanskaya, Itziar Laka & Marina Nespor. 2017. Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory and Cognition* 45(5). 863–876. DOI: 10.3758/s13421-017-0700-9.

Ots, Nele & Piia Taremaa. 2022. A perceptual study of language chunking in Estonian. *Open Linguistics* 8(1). 1–26. DOI: 10.1515/opli-2020-0182.

Peters, Benno. 1999. *Prototypische Intonationsmuster in deutscher Lese- und Spontansprache.* Magister Thesis, Universität Kiel. 1–171. https://www.ipds.uni-kiel.de/kjk/pub_exx/bp1999_1/aipuk34_bp-ma.pdf.

Petrone, Caterina, Hubert Truckenbrodt, Caroline Wellmann, Julia Holzgrefe-Lang, Isabell Wartenburger & Barbara Höhle. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61. 71–92. DOI: 10.1016/j.wocn.2017.01.002.

Pierrehumbert, Janet B. 1979. The perception of fundamental frequency declination. *Journal of Acoustical Society of America* 66(2). 363–369. DOI: 10.1121/1.383670.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation.* Massachusetts Institute of Technology. (Doctoral dissertation). http://dspace.mit.edu/handle/1721.1/16065.

Pierrehumbert, Janet B. & Julia Bell Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 271–312. Cambridge, MA: MIT Press. DOI: 10.7916/D8KD24FP.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Riesberg, Sonja, Janina Kalbertodt, Stefan Baumann & Nikolaus P. Himmelmann. 2020. Using Rapid Prosody Transcription to probe little-known prosodic systems: The case of Papuan Malay. *Laboratory Phonology* 11(1)(8). 1–35. DOI: 10.5334/labphon.192.

Schild, Ulrike, Angelika B. C. Becker & Claudia K. Friedrich. 2014. Phoneme-free prosodic representations are involved in pre-lexical and lexical neurobiological mechanisms underlying spoken word processing. *Brain and Language* 136(100). 31–43. DOI: 10.1016/j.bandl.2014.07.006.

Schubö, Fabian & Sabine Zerbian. 2023. The patterns of pre-boundary lengthening in German. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell

Wartenburger (eds.), *Prosodic boundary phenomena*, 1–34. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777526.

Schuetze-Coburn, Stephan, Marian Shapley & Elizabeth G. Weber. 1991. Units of intonation in discourse: A comparison of acoustic and auditory analyses. *Language and Speech* 34(3). 207–234. DOI: 10.1177/002383099103400301.

Silbert, Lauren J., Christopher J. Honey, Erez Simony, David Poeppel & Uri Hasson. 2014. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences USA* 111(43). E4687–E4696. DOI: 10.1073/pnas.1323812111.

Simon, Anne Catherine & George Christodoulides. 2016. Perception of prosodic boundaries by naïve listeners in French. In *Proceedings of Speech Prosody*, 1158–1162. DOI: 10.21437/SpeechProsody.2016-238.

Thorsen, Nina Grønnum. 1985. Intonation and text in Standard Danish. *The Journal of the Acoustical Society of America* 77(3). 1205–1216. DOI: 10.1121/1.392187.

Trouvain, Jürgen, William J. Barry, Claus Nielsen & Ove Kjeld Andersen. 1998. Implications of energy declination for speech synthesis. In *Proceedings of the 3rd ESCA/COCOSDA workshop on speech synthesis*, 47–52.

Ulbrich, Christiane. 2002. A comparative study of intonation in three standard varieties of German. In *Proceedings of Speech Prosody*, 671–674. https://www.isca-speech.org/archive/pdfs/speechprosody_2002/ulbrich02_speechprosody.pdf.

Wagner, Michael & Michael McAuliffe. 2019. The effect of focus prominence on phrasing. *Journal of Phonetics* 77. 100930. DOI: 10.1016/j.wocn.2019.100930.

Wellmann, Caroline, Julia Holzgrefe-Lang, Hubert Truckenbrodt, Isabell Wartenburger & Barbara Höhle. 2023. Developmental changes in prosodic boundary cue perception in German-learning infants. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 119–156. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777532.

White, Laurence, Silvia Benavides-Varela & Katalin Mády. 2020. Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? *Journal of Phonetics* 81. 10098. DOI: 10.1016/j.wocn.2020.100982.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America* 91(3). 1707–1717. DOI: 10.1121/1.402450.

Yuan, Jiahong & Mark Y. Liberman. 2014. F0 declination in English and Mandarin broadcast news speech. *Speech Communication* 65. 67–74. DOI: 10.1016/j.specom.2014.06.001.

Chapter 4

Developmental changes in prosodic boundary cue perception in German-learning infants

Caroline Wellmann[a], Julia Holzgrefe-Lang[a], Hubert Truckenbrodt[b], Isabell Wartenburger[a] & Barbara Höhle[a]

[a]University of Potsdam [b]Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin

Previous investigations suggest that the main prosodic cues characterizing intonation phrase boundaries (IPBs), namely pitch change, final lengthening, and pause, have different weightings in perception. Such a weighting of IPB cues may be subject to crosslinguistic variation and seems to develop during the first year of life. For German, eight-month-old infants were found to detect an IPB signaled by pitch change combined with final lengthening in a behavioral discrimination task, even in the absence of a pause (Wellmann et al. 2012). Assessing the developmental course of prosodic boundary detection, the present study tested six-month-old German-learning infants with the same discrimination task in the headturn preference paradigm. Stimuli were presented in two different prosodic groupings, as a sequence either without or with an internal boundary after the second name, *[Moni und Lilli und Manu]$_{IP}$* vs. *[Moni und Lilli]$_{IP}$ [und Manu]$_{IP}$*. The internal IPB was systematically varied with respect to the amount and combination of cues. Infants were familiarized to sequences without an IPB and then tested on their discrimination of both prosodic groupings. We found that infants detected the boundary when it was cued by all cues (Exp. 1) and by pause and lengthening (Exp. 3). However, when the IPB was only marked by a combination of pitch and lengthening, they failed (Exp. 2), even when familiarization duration was doubled (Exp. 2a). This points to a crucial role of the pause cue at six months. Taken together with previous results, our data suggest a development towards an adult-like boundary perception that no longer requires the pause cue between six and eight months. We argue that this behavioral change reflects a shift of attention to boundary markings that are functionally relevant in the ambient language.

Caroline Wellmann, Julia Holzgrefe-Lang, Hubert Truckenbrodt, Isabell Wartenburger & Barbara Höhle. 2023. Developmental changes in prosodic boundary cue perception in German-learning infants. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 119–156. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777532

1 Introduction

During their first year of life infants pass through a phase of perceptual reorganization, in which their speech perception is sharpened for acoustic properties that are functional in the language they are exposed to but attenuates for acoustic properties that are not (for a review, see Maurer & Werker 2014). Perceptual reorganization was initially shown for vowels and consonants, with numerous findings suggesting an increasing ability to discriminate native sound contrasts but decreasing performance with non-native sounds (for vowels, e.g., Kuhl et al. 1992, Polka & Werker 1994; for consonants, e.g., Best & McRoberts 2003, Kuhl et al. 2006, Werker & Tees 1984). More recently, perceptual reorganization has also been shown for suprasegmental prosodic aspects of language like lexical tone (Götz et al. 2018, Mattock & Burnham 2006, Mattock et al. 2008, Yeung et al. 2013) and lexical stress (Bijeljac-Babic et al. 2012, Höhle et al. 2009, Jusczyk et al. 1993, Skoruppa et al. 2009).

The present paper deals with a potential developmental change in a further area of prosody, namely in the perception of acoustic cues that mark major prosodic boundaries, specifically boundaries at the edges of Intonation Phrases (IP, Nespor & Vogel 1986, Selkirk 1986). Three main cues are considered to mark these boundaries across languages: a change in fundamental frequency (F0) in terms of boundary tones or a pitch reset, a lengthening of preboundary segments, and the insertion of a pause (e.g., Hirst & Di Cristo 1998, Nespor & Vogel 1986, Price et al. 1991, Vaissière 1983).

The edges of IPBs usually coincide with syntactic clause boundaries (Selkirk 2005; for German: Truckenbrodt 2005), and infants as well as adults benefit from this close syntax-prosody mapping. Over the past thirty years, multiple studies have provided consistent evidence that infants are highly sensitive to prosodic boundary information and use it to segment the continuous speech signal into linguistically relevant units (Gout et al. 2004, Hirsh-Pasek et al. 1987, Kemler Nelson et al. 1989, Nazzi et al. 2000, Schmitz 2008, Shukla et al. 2011). Assuming that each prosodic boundary cue contributes individually to boundary perception, previous research has focused on the specific roles of pitch change, final lengthening, and pause in adult as well as infant listeners (e.g., Aasland & Baum 2003, Johnson & Seidl 2008, Lehiste et al. 1976, Petrone et al. 2017, Sanderman & Collier 1997, Scott 1982, Seidl 2007, Streeter 1978, Zhang 2012).

Studies with American English-learning infants found evidence for a developmental change in perception of these boundary cues (Seidl 2007, Seidl & Cristià 2008). In a series of experiments with varying cue manipulations, Seidl (2007) tested six-month-old American English-learning infants, applying an experimental design based on Nazzi et al. (2000). In a familiarization phase, infants were presented with two sequences of the same words extracted from two different naturally recorded text passages of infant-directed speech. One familiarization sequence was a complete clause (within clause stimuli), e.g., *# Leafy vegetables taste so good #*. The other one contained an internal clause boundary and thus a major prosodic boundary (non-clause stimuli), e.g., *leafy vegetables # taste so good*. During the test phase, the original text passages from which the familiarization sequences had been extracted were presented. In the first experiment, the internal boundary of the non-clausal familiarization sequence was fully marked by all three boundary cues, that is, a pitch reset at the juncture, final lengthening, and a pause. Results indicated that infants were better able to recognize the clausal than the non-clausal sequence in the continuous speech presented during the test phase. In subsequent experiments, the familiarization sequences were varied by acoustic manipulation of one or two of the three essential prosodic cues to find out if clause segmentation would still be possible. When either pause or final lengthening was neutralized, infants still preferred the passage with the familiar clausal sequence. Consequently, neither of these two cues was necessary to evoke the detection of a prosodic boundary. However, when the pitch cue was neutralized, infants no longer showed a preference for the clausal sequence. Yet pitch alone was not a sufficient marker: when it was present as a single cue with neutralized pause and final lengthening, clause segmentation was not successful. Hence, Seidl (2007) concluded that American English-learning six-month-old infants need the pitch cue in combination either with pause or with final lengthening for clause segmentation.

Seidl & Cristià (2008) extended these investigations to four-month-old American English-learning infants. Tested with the same experimental material and design, the younger group was only successful in clause segmentation when all three cues in combination signaled the boundary. The authors concluded that four-month-old infants' clause segmentation is based on a holistic processing that relies on a coalition of all boundary cues while six-month-old American English-learning infants are already able to process the cues independently of each other. Crosslinguistic evidence provides first hints that boundary perception by six-month-olds is modulated by specific prosodic properties of the language that the infants are learning. Johnson & Seidl (2008) used the same experimental procedure as Seidl (2007) to test Dutch-learning six-month-olds with Dutch materials. In contrast to their American English age-mates, the Dutch infants only showed evidence for clause segmentation when the prosodic boundaries were marked by the combination of pitch, lengthening, and a pause, but not when the pause was removed from the materials. The authors argue that pauses may be a strong marker of prosodic boundaries in Dutch and thus important for infants' boundary detection. In fact, a comparison of the speech materials that were naturally recorded for the American English and the Dutch experiments showed that in the Dutch passages the pauses at the clause boundaries were twice as long as in the American English ones while the American English passages showed a much larger pitch reset after the juncture than the Dutch materials.

Broadening the crosslinguistic perspective, Wellmann et al. (2012) investigated German-learning infants on their perception of boundary cues in bracketed lists of names as *[A and B and C]* in contrast to *[A and B] [and C]* indicating either one group of three people or a group of two and a single person. According to Petrone et al. (2017), German speakers typically employ the IP to signal such a grouping (see also Huttenlauch et al. 2023 [this volume]). Indeed, analyzing the respective cues at the natural internal boundary in Wellmann et al.'s materials revealed typical characteristics of an IPB: pitch changes, specifically an upstep on the second peak and a partial pitch reset, a lengthening of the preboundary vowel, and the employment of a pause (for similar findings concerning prosodic boundary cues in German, see also Féry & Kentner 2010, Schubö & Zerbian 2023 [this volume], Truckenbrodt 2007a, 2016). Hence, the sequences either formed a single IP, *[Moni und Lilli und Manu]*$_{IP}$ or were made up of two IPs with an internal IPB after the second name, *[Moni und Lilli]*$_{IP}$ *[und Manu]*$_{IP}$. The internal IPB was the focus of our investigations on prosodic cue perception.

Unlike previous studies (Johnson & Seidl 2008, Nazzi et al. 2000, Seidl 2007), Wellmann et al. (2012) and the present study tested the detection of the prosodic boundary not by a clause segmentation task, but by a discrimination task. In this discrimination task, two groups of eight-month-old infants were familiarized with a sequence of one prosodic type, either with or without an internal IPB. In the subsequent test phase, all infants were presented with sequences of both prosodic types to test whether they discriminated between them. Given that previous research on infants' attunement to features of segmental phonology also used discrimination tasks (Mattock & Burnham 2006, Mattock et al. 2008, Polka & Werker 1994, Werker & Tees 1984), we assumed that such a discrimination task should be suited to reveal differences in prosodic boundary information as well. We considered the methodology as an important contribution since the same materials were suitable for use in a behavioral study with adults, as well as for use in ERP studies with adults and infants (Holzgrefe-Lang et al. 2016, 2018). Moreover, the material can in principle be used in other languages as well (for French, van Ommen et al. 2020).

Wellmann et al. (2012)'s results revealed that eight-month-old infants preferred to listen to sequences of the new prosodic grouping (i.e., sequences with an IPB) after familiarization with sequences without an IPB. This indicated successful discrimination of the prosodic patterns. In subsequent experiments the prosodic boundary information was systematically varied by adding a cue or a subset of cues to the original sequence without an internal IPB after the second name. When the IPB was signaled by a pitch rise and final lengthening in combination but without a pause, eight-month-olds still successfully discriminated the sequences with boundary cues from the sequences without an internal IPB. However, when the IPB was signaled solely by a pitch rise or by final lengthening, infants did not discriminate the two prosodic conditions. These findings suggest that pitch change and lengthening in combination, but not as single cues, are sufficient for IPB detection in eight-month-old German-learning infants while the presence of a pause is not necessary.

Interestingly, the discrimination pattern of the German eight-month-olds mirrored a pattern that Holzgrefe-Lang et al. (2016) observed in a prosodic judgment task using the same stimuli with German-speaking adults. In this task, participants judged via button-press whether the stimuli contained an internal boundary or not. The results revealed that stimuli containing a pitch change and final lengthening in combination but no pause were judged as sequences with an IPB. In contrast, stimuli that contained only a pitch change were predominantly judged as sequences without an internal boundary, and sequences with only lengthening were judged at chance level, indicating no categorization.

Although the two tasks – the discrimination task with infants and the prosodic judgment task with adults – may place some different requirements on the participants and each group's data was analyzed on its own, the similarity in the results across the two studies indicates that German-learning eight-month-olds' sensitivity to prosodic boundary cues already resembles that of German adults. The question arises whether the discrimination pattern found in the German eight-month-olds is in fact the result of a perceptual attunement from a solely acoustically driven perception based on the presence of the salient pause cue to a more sophisticated linguistically affected perception relying on pitch change and lengthening.

Therefore, six-month-old German-learning infants were studied with the same experimental paradigm and the same stimuli as used by Wellmann et al. (2012); that is, sequences without an internal IPB and sequences with an internal boundary cued by the full set or a subset of pitch, lengthening, and pause cues had to be discriminated. In Experiment 1, detection of a prosodic boundary that is fully marked by the combination of all naturally occurring cues was investigated.

We assumed that infants are able to detect this boundary, as previous research presenting infants with artificial pauses at boundary and non-boundary locations has revealed that German infants are highly sensitive to the correlation of prosodic boundary information already in their first half year of life (Schmitz 2008). Experiment 2 examined whether pitch change and final lengthening in combination are sufficient boundary markers or whether pause is a necessary cue for this age group. Here we aimed to clarify whether prosodic boundary detection at six months already reflects an attunement towards linguistically relevant markings or whether it is rather influenced by the perceptual salience of cues. In Experiment 2a we used the same stimuli as in Experiment 2, but with a prolonged familiarization. We hypothesized that infants develop a sensitivity towards boundaries that are not cued by pause between six and eight months. In Experiment 2a, we asked whether this sensitivity would show up already at six months under optimized experimental conditions, that is, after a prolonged exposure. We hypothesized that the double amount of presentations of the familiarization sequence might lead to a more stable mental representation and would thus release (working) memory capacity to thoroughly explore the new stimulus with its differences. In Experiment 3 we investigated whether the combination of pause and final lengthening provides sufficient information for boundary detection or whether – as has been shown for younger American English-learning infants – only a combination of all cues would evoke boundary detection. In the following, we will successively introduce each experiment with its participants, stimuli, procedure, and its descriptive results. Subsequently, a statistical analysis across all four experiments will be reported, followed by a general discussion.

2 Experiment 1: The influence of pitch, final lengthening, and pause

Experiment 1 tested whether German-learning six-month-old infants are able to perceive an IPB that is signaled by the three main prosodic cues pitch change, final lengthening, and pause.

2.1 Participants

A group of twenty-four six-month-old infants (12 girls, 12 boys) was tested. Their mean age was 6 months, 11 days (range: 6 months, 2 days to 6 months, 27 days). Nine additional infants were tested but not included in the data for the following reasons: failure to complete the experiment (1), crying or fussiness (4), mean listening times of less than 3 s per condition (1), technical problems (2), and noise in the surroundings due to construction work (1).

All infants who participated in this and the following experiments were from monolingual German-speaking families, born full-term, and with normal hearing. They were recruited from birth lists obtained through the Potsdam city hall archives. All parents signed informed consent. None of the infants tested in the present study participated in more than one experiment.

2.2 Stimuli

All stimuli used in Experiments 1 and 2 were identical to those that were presented to eight-month-olds in the study by Wellmann et al. (2012): the stimuli consisted of a sequence of three German names containing only sonorant sounds *(Moni, Lilli, Manu)*, which allowed a reliable measure of F0 and were suitable for acoustic manipulation. The names were coordinated by *und* (‘and’). A young female adult, a German native speaker from the Brandenburg area, was instructed to read the sequence in two different prosodic groupings indicated by different bracketing:

(1) (Moni und Lilli und Manu) – without internal IPB

(2) (Moni und Lilli) und Manu – with internal grouping

Both sequences contained the same string of names and differed only in grouping either all three names together as shown in (1) or grouping the first two names together and the final one apart as shown in (2). Sequences of type (1) were produced as a single IP, without an internal boundary. In contrast, sequences of type (2) consisted of two IPs, with an internal IPB after the second name. The speaker repeated each sequence six times, resulting in six recordings per prosodic type. The intended grouping was confirmed by two independent listeners who were naïve with respect to the given bracketing. Recordings were made in an anechoic chamber equipped with an Audio-Technica AT4033A studio microphone, using a C-Media Wave soundcard at a sampling rate of 22,050 Hz with 16-bit resolution. Examples of both kinds of prosodic phrasing are depicted in Figure 1A and B.

The acoustic analysis of the recordings revealed clear acoustic differences between the two prosodic phrasings on and after the second name (see Table 1).

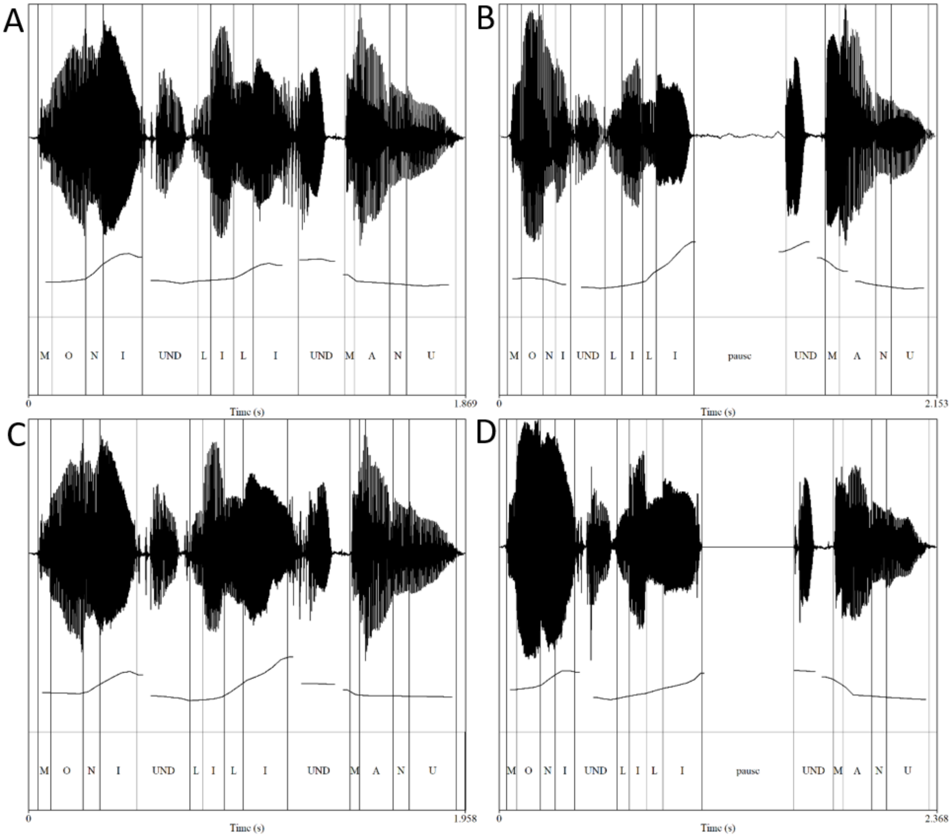


Figure 1: Oscillograms and pitch contours aligned to the text. Vertical lines mark the segmental boundaries. The hash mark indicates the silent pause after the IPB. (A) Sequence without an IPB used in Exp. 1, (B) Sequence with a fully marked internal IPB used in Exp. 1, (C) Sequence with pitch change and final lengthening used in Exp. 2 and 2a, (D) Sequence with pause and final lengthening used in Exp. 3.

Table 1: Mean values (range) of the acoustic correlates of prosodic boundary in the six experimental sequences *Moni und Lilli und Manu* without and with an internal IPB, respectively. BT: boundary tone (Hz), maxF0 on Name2's final vowel; PR: pitch rise (Hz), maxF0 on Name2 minus minF0 on Name2; PPR: partial pitch reset (Hz), maxF0 on Name2 minus minF0 on Name2; FL: final lengthening (ms), duration of Name2's final vowel; P: pause (ms), duration of pause after Name2. †: in semitones.

	Boundary cue				
	BT	PR†	PPR†	FL	P
Without internal IPB					
Mean	277	88 (6.7)	5×rise: −12 (−0.7) 1×fall: 18 (1.1)	99	0
SD	264–293	77–110 (5.8–8.2)	−23 – −5 (−1.3 – −0.3)	91–110	0
With internal IPB					
Mean	397	220 (14.0)	6×fall: 55 (2.5)	175	506
SD	371–422	197–240 (12.8–14.6)	36–96 (1.7–4.4)	162–186	452–556

2.2.1 Preboundary pitch movement

Sequences without an internal IPB that form one single IP were characterized by F0 lowering, with an accentual pitch rise on the first name, followed by a smaller pitch rise on the second name, that is, a downstep pattern (Truckenbrodt 2007b). Sequences with an internal IPB exhibited a flat tonal contour (plateau) on the first name, followed by a large pitch rise on the second name, starting at the second syllable, and leading to an upstepped peak, a high boundary tone, at the final vowel. This pitch rise on the second name (measured as the difference between the maximum and minimum pitch on the second name in semitones) was 2.5 times greater and led to a higher maximum pitch than the small rise occurring at the same location in sequences without an IPB. Hence, the F0 contour of sequences with an internal IPB clearly indicated the following prosodic boundary. This tonal contour resembled the most common realization of internal IPBs in similar German sequences of names investigated in Petrone et al. (2017).

2.2.2 Postboundary pitch movement

Postboundary pitch reset was measured as the difference between the maximum pitch on the final vowel of the second name and the maximum pitch on the vowel of the conjunction. In sequences without an internal IPB there was no relevant pitch difference: in five of the six recordings the height of the downstepped second peak was slightly higher at the conjunction (on average by 0.7 semitones), whereas in one recording it was slightly lower (by 1.1 semitones). In sequences with an internal IPB a partial pitch reset, one step below the preboundary upstep, occurred (see Truckenbrodt 2007b for a similar partial reset). This was expressed by a pitch fall of 2.5 semitones on average.

2.2.3 Final lengthening

To explore final lengthening, the duration of the second name's final vowel [i] was compared in sequences with and without an internal IPB. Its duration was 1.8 times longer in the grouping with an IPB, indicating a strong lengthening cue (cf. Kohler 1983).

2.2.4 Pause duration

Finally, the duration of the silent interval after the second name was measured. A pause with an average duration of 506 ms occurred in sequences with an internal IPB, whereas no pause was present in sequences without an internal IPB.

Taken together, in sequences with an internal IPB all acoustic correlates of the three main prosodic boundary cues were observed: a change in F0 (mainly, a preboundary pitch rise), a lengthening of the preboundary vowel, and the occurrence of a pause. The recorded sequences were used to create sound files for presentation during the experiment. All recordings were scaled to a mean intensity of 70 dB. For each prosodic type, the six recordings were randomly concatenated with a silent interval of 1 s inserted between them. In this way, six sound files per prosodic grouping were created such that each file consisted of a different order.

Due to the missing durational cues (final lengthening and pause), sequences without an internal IPB were shorter than those with an internal IPB. The average duration of sequences without an internal IPB was 1.76 s (range: 1.67–1.87 s), while it was 2.16 s (range: 2.13–2.2 s) for sequences with an internal IPB. To match the sound files of the two prosodic types with respect to overall length, the number of sequences within each file was varied. As a result, sound files of the condition with an internal IPB contained six sequences and had an average duration of 18.97 s, and sound files of the condition without an internal IPB contained seven sequences (i.e., one random recording was repeated), leading to an average duration of 19.32 s (range: 19.16–19.43 s). The difference in the number of sequences was crucial in order to present sound files of similar lengths during the experimental trials.

2.3 Procedure

In all experiments presented here, infants were tested using the headturn preference procedure (HPP) including a familiarization phase. During the experiment, the infant was seated on the lap of a caregiver in the center of a test booth. Inside this booth three lamps were fixed: a green one on the center wall, and a red one on each of the side walls. Directly above the green lamp was an opening for the lens of a video camera. Behind each of the red lights a JBL Control One loudspeaker was mounted. Each experimental trial started with the blinking of the green center lamp. When the infant oriented to the green lamp, it was turned off and one of the red lamps on a side wall started to blink. When the infant turned her head towards the red lamp, the speech stimulus was started, delivered via a Sony TA-F261R audio amplifier to the loudspeaker on the same side. The trial ended when the infant turned her head away for more than 2 s, or when the end of the speech file was reached. If the infant turned away for less than 2 s, the presentation of the speech file continued but the time spent looking away was not included in the total listening time. The whole session was digitally videotaped. The experimenter's coding was recorded and served for the calculation of the duration of the infant's head turns during the experimental trials. The caregiver listened to music over headphones to prevent influences on the infant's behavior. Furthermore, she was instructed not to interfere with the infant's behavior during the experiment. The experimenter sat in an adjacent room, where she observed the infant's behavior on a mute video monitor and controlled the presentation of the visual and the acoustic signals by a button box. The experimenter was blind with respect to the type of acoustic stimuli presented during familiarization and testing.

An experimental session consisted of a familiarization phase immediately followed by a test phase. For Experiment 1, we familiarized half of the infants with sequences without an internal IPB, while the other half listened to sequences with an internal IPB. For both groups, familiarization lasted until at least 20 sequences were presented. Given that sequences without an IPB were shorter than sequences with an internal IPB, familiarization timing differed slightly. That is, when infants were familiarized with sequences without an internal IPB, familiarization lasted until 55 s of listening time had been accumulated. When familiarized with sequences with an internal IPB, infants had to accumulate 63 s of listening time.

After familiarization, infants immediately passed through the test phase that comprised twelve trials. Half of the test trials contained the identical sound files previously presented during familiarization (familiar test trials). The other six trials contained the sound files of the other prosodic grouping (novel test trials). The test trials were grouped in three blocks of four trials each. Two out of these four trials contained sequences with an internal IPB, the others contained sequences without an internal IPB. Within each block, test trials were randomly ordered with the side of presentation being counterbalanced for each prosodic type. Based on the infant's head turns the listening time to each test trial was measured. The duration of each experimental session varied between four and six minutes, depending on the infant's behavior.

2.4 Descriptive results

We analyzed the data for each familiarization group on a descriptive level. Within this sample we observed clear numerical differences in the listening times. The group of infants that was familiarized with sequences with an internal IPB listened on average for 7.69 s (SD = 3.41 s) to novel test trials and for 7.52 s (SD = 2.90 s) to familiar test trials; that is, the mean listening time between novel and familiar test trials only differed by $M_{Diff} = 0.17$ s. Six out of twelve infants listened longer to the novel test trials.

The group of infants that was familiarized with sequences without an internal IPB listened on average for 9.65 s (SD = 2.96 s) to novel test trials and for 8.06 s (SD = 3.02 s) to the familiar ones (see Figure 2). Hence, the familiarization with sequences without an internal IPB yielded a novelty preference with a mean listening time difference of $M_{Diff} = 1.58$ s. Nine out of twelve infants listened longer to the novel test trials.

Overall, a clearer numerical difference between the two prosodic patterns showed up after familiarization with sequences without internal IPB, whereas the familiarization with sequences with an internal IPB seemed to be much less or even not effective, as also evidenced in other studies (see van Ommen et al. 2020 and Wellmann et al. 2012 for a discussion of this asymmetric behavior). Given the constraints we usually encounter in infant research (small sample sizes, high drop-out rates related to infant behavior) we therefore decided to run only the familiarization with sequences without an internal IPB in the subsequent experiments of the present study.[1]

3 Experiment 2: The influence of pitch and final lengthening

Experiment 2 examined whether a subset of prosodic cues would suffice to trigger the perception of a boundary in six-month-old infants. Specifically, the impact of the combination of a rising pitch contour and final lengthening was under focus, questioning the necessity of the pause cue.

3.1 Participants

Sixteen infants (8 girls, 8 boys) were tested. The mean age was 6 months, 14 days (range: 5 months, 28 days to 6 months, 29 days). Four additional infants were tested but not included in the data analysis for the following reasons: failure to complete the experiment (1), crying or fussiness (1), and mean listening times of less than 3 s per condition (2).

---

[1] With the same Experiment 1, we also tested 24 four-month-old infants ($n = 12$ in each familiarization group) in a slightly modified HPP setup (to adapt for the limited head movements at that age the position of the side lamps was moved to the edges of the front wall). Four-month-old infants that were familiarized to sequences without an IPB had mean listening times of 10.5 s (SD = 3.59 s) to novel test trials and 10.25 s (SD = 3.26 s) to the familiar ones, $t(11) = 0.483$, $p = 0.639$, two-tailed. The group familiarized to sequences with an IPB listened on average 10.01 s (SD = 4.12 s) to the novel test condition and 10.67 s (SD = 4.90 s) to the familiar one, $t(11) = -1.1$, $p = 0.295$, two-tailed. Given the null result we did not continue in testing four-month-olds.

3.2 Stimuli

Experiment 2 involved stimuli of two prosodic types: One condition comprised the same six sequences without an internal IPB as in Experiment 1. The other one consisted of six sequences with only pitch rise and lengthening cues indicating the boundary. For this prosodic type, sequences without an internal IPB were locally acoustically manipulated with respect to F0 on the second name and the duration of its final vowel. A specific pitch reset cue, that is, a manipulation of F0 at the position of the postboundary conjunction, was not implemented, since in the stimuli of Experiment 1 the postboundary peak was utterance-final and generally low. Stimuli manipulations were carried out the same way as in Wellmann et al. (2012).

The stimuli without a boundary were selected as the basis for the acoustic manipulations to avoid a potential influence of additional cues that may contribute to IPB marking and perception. Thus, the crucial boundary information, here a rising pitch contour and final lengthening, was added to the sequences without an internal IPB. Hence, experimental effects can clearly be attributed to the acoustic properties under investigation. By using these local cue manipulations the stimuli with IPB cues differed from sequences that were used in the condition without an internal IPB only within a predefined critical region and by controlled acoustic properties. However, this local manipulation led to the concession that sequences with inserted pitch rise and lengthening differed from natural sequences with an internal IPB with respect to the pitch contour of the first name. The original recordings without an internal IPB had an accentual peak on the first name, that is, a pitch cue to a phonological phrase boundary (Figure 1C). The accentual peak on the first name was always lower than the peak of the H% in the second name (M_{F0MAX} = 317 Hz vs. 388 Hz). However, this kind of cue was not present in the naturally produced sequences with an internal IPB, but was preserved in sequences with inserted cues since sequences without an internal IPB were the base for sequences with inserted cues and cue manipulations were restricted to the second name. Hence, the inserted pitch rise was preceded by another smaller pitch cue. If the pitch rise on the second name is parsed globally, that is, in relation to the previous pitch contour, the pitch rise cue in the manipulated sequences is less pronounced and potentially less salient than the pitch rise cue in natural sequences with an internal IPB.

The acoustic manipulation was carried out with Praat (Boersma & Weenink 2019). As the phonetic magnitude of prosodic cues differs across languages, and also within a language depending on the syntactic structure (e.g., for pausing in German, see Butcher 1981), there is no unique value for each prosodic cue.

Hence, we decided to implement the same phonetic magnitude for each cue that was present in the corresponding naturally produced stimulus with an internal IPB, proceeding in the same way as other studies that have employed cue manipulations (e.g., Seidl 2007).

Manipulation steps were the following: To implement the pitch rise, first, the pitch contour of the sequences without an internal IPB was stylized (two semitones). This transformation decreases pitch perturbations by reducing the number of pitch points. Second, the pitch points on the second name were set to the reference values. The reference values of F0 were measured on the second name in the six original sequences with an internal IPB (used in Experiment 1), namely at the midpoints of the four segments [l], [ɪ], [l], [i] and at the position of the maximum pitch present on the final vowel. For the manipulation of the pitch contour, pitch points with the mean values at these time points (176 Hz, 183 Hz, 224 Hz, 305 Hz, and 397 Hz) were inserted into the stylized sequences without an internal IPB at the same positions. After PSOLA resynthesis in Praat, the six new stimuli contained a natural sounding pitch rise of 212 Hz (13.65 semitones) leading to an H% with a mean value of 388 Hz. To implement final lengthening, the final vowel [i] of the second name was lengthened to 180%. This factor was chosen because in the natural stimuli, the crucial vowel was on average 1.8 times longer in sequences with an internal IPB than in sequences without an internal IPB (Table 1). A sequence with manipulated pitch and lengthening is depicted in Figure 1C.

To avoid comparing natural with acoustically manipulated material, we carried out a slight acoustic manipulation in sequences without an internal IPB as well, that is, the stylization of the pitch contour (two semitones). After pitch stylization, sequences were resynthesized using the PSOLA function.

Sequences without an internal IPB lasted on average 1.76 s (range: 1.67–1.87 s), while sequences with inserted pitch and lengthening had a mean duration of 1.84 s (range: 1.74–1.96 s). From these sequences six differently ordered sound files per prosodic type were created to be used as experimental trials. The interstimulus interval between the sequences and within a sound file was 1 s. All sound files contained seven sequences (one random recording was repeated). The files containing sequences without an internal IPB had an average duration of 18.33 s (range: 18.23–18.43 s) and the files containing sequences with inserted pitch and lengthening cues lasted on average 18.81 s (range: 18.79–19.01 s).

3.3 Procedure

All infants were familiarized with sequences without an internal IPB. The familiarization lasted until at least 20 sequences had been presented, resulting in a minimum of 52 s of accumulated listening time. The familiarization was immediately followed by a test phase with twelve trials. As in Experiment 1, half of the test trials contained the same sound files that the infants had heard during familiarization. The other half contained the files of the sequences with pitch and lengthening cues, which had not been presented during familiarization. All twelve test trials were grouped in three blocks of four trials each (two of each prosodic type in a random order). The infant's listening time to each test trial was measured.

3.4 Descriptive results

Infants tested in Experiment 2 showed a mean listening time of 7.09 s (SD = 2.1 s) to the novel test trials and a mean listening time of 7.20 s (SD = 2.43 s) to the familiar test trials (see Figure 2). Eight out of 16 infants had longer listening times to the familiar test trials.

4 Experiment 2a: The influence of pitch and final lengthening after prolonged familiarization

To verify that the non-discrimination in Experiment 2 was due to the composition of the stimuli, we modified the experimental design by doubling the familiarization time. Considering a longer familiarization to enable successful discrimination stems from findings of studies that tested French-learning infants' discrimination of rhythmic patterns (Bijeljac-Babic et al. 2012, Höhle et al. 2009, Skoruppa et al. 2009). For French, a language without contrastive stress at the word level, the perception of prosodic cues indicating lexical stress has been shown to be hard for infant learners and adult listeners (Bhatara et al. 2013, Höhle et al. 2009). In a study by Bijeljac-Babic et al. (2012) monolingual ten-month-old French-learning infants exhibited a null result in discriminating an iambic and a trochaic version of a pseudo-word after a one-minute familiarization. However, when familiarization duration was increased to two minutes, they were successful at discriminating the stress patterns as indicated by a novelty effect. Bijeljac-Babic et al. concluded that the null result after the short familiarization could not be interpreted as a general inability to distinguish the two stress patterns, but was due to the short familiarization. Regarding the novelty effect after long familiarization, they drew on the model by Hunter & Ames (1988) that would predict novelty preferences in relatively easy discrimination conditions.

In our case of discriminating lists of names with and without IPB cues, it is important to consider that boundary perception without the pause cue is successful in older infants. This sensitivity seems to arise between six and eight months. If six-month-old German-learning infants are already at the beginning of this development, in the light of the Hunter and Ames model, discrimination might show up with reduced task difficulty. Following Bijeljac-Babic et al. (2012) we hypothesized that a more robust mental representation of the stimuli presented during familiarization may improve the ability to detect differences between the familiar and the novel stimuli. In the following experiment, we therefore doubled the amount of presentations of the familiarization stimulus in order to help infants building up a more robust mental representation of the sequences without an internal IPB. Through this modification, infants might be able to accomplish a still difficult task for their age such as the detection of a boundary signaled only by pitch and lengthening cues.

4.1 Participants

Twenty-three six-month-old infants (12 girls, 11 boys) were tested. The mean age was 6 months, 15 days (range: 6 months, 0 days to 6 months, 26 days). All infants were from monolingual German-speaking families, born full-term and normal-hearing. Thirty-one additional infants were tested but their data were not included in the analysis for the following reasons: failure to complete the experiment (6), crying or fussiness (15), mean listening times of less than 3 seconds per condition (4), technical problems (2), experimenter error (2), parental interference (1) and outlying listening times due to steady fixation (1). Drop-out rate was especially high, primarily due to infants' fussiness and failure to finish the experiment (accounting for 68% of all drop-outs). The longer lasting familiarization which increased the total duration of the experiment to about 6 to 10 minutes (in contrast to six minutes with the original familiarization duration) may have reduced infants' attention.

4.2 Stimuli

Stimuli were exactly the same as in Experiment 2.

4.3 Procedure

Infants were familiarized with sequences without an internal IPB. The familiarization duration was set to 104 s. After familiarization, infants listened to exactly the same twelve test trials as used in Experiment 2, half of them being sequences with inserted pitch and lengthening cues, the other half sequences without an internal IPB.

4.4 Descriptive results

Infants tested in Experiment 2a showed a mean listening time of 6.86 s (SD = 2.01 s) to the novel test trials, and a listening time of 7.52 s (SD = 2.25 s) to the familiar test trials (see Figure 2). Fifteen out of 23 infants listened longer to familiar test trials.

5 Experiment 3: The influence of pause and final lengthening

In the following Experiment 3 the boundary was cued by a pause in combination with final lengthening, but without any pitch cue. The aim of this experiment was to investigate whether six-month-olds would respond to a boundary that is cued by a subset of the naturally occurring cues including a pause. The combination of pause and lengthening was chosen because a pause rarely occurs as the only cue in German (only at 1.3% of all boundaries in the analysis by Peters et al. 2005), and would thus sound unnatural as the only inserted cue[2], and because the combination of pause with final lengthening occurs more frequently (8.4%) in spoken German than the combination of pause and pitch (4.9%, values by Peters et al. 2005). Moreover, this combination is interesting to look at crosslinguistically, as six-month-old American English-learning infants failed to perceive a boundary signaled only by the combination of pause and lengthening cues (Seidl 2007).

5.1 Participants

Sixteen infants (8 girls, 8 boys) were tested. The mean age was 6 months, 10 days (range: 5 months, 14 days to 6 months, 28 days). Eleven additional infants were tested but not included in the data for the following reasons: failure to complete the experiment (1), crying or fussiness (6), mean listening times of less than 3 s per condition (2), and technical problems (2).

---

[2] We are grateful to one reviewer who raised the question whether discrimination would be possible with pause as the only boundary cue. We hypothesize that a similar experiment with a boundary cued by pause only would lead to successful discrimination as well. This hypothesis is based on infants' successful discrimination between clauses with artificially inserted pauses at non-boundary locations and clauses with pauses at natural boundary positions with co-occurring boundary cues (Schmitz 2008).

5.2 Stimuli

In Experiment 3, we contrasted sequences without an internal IPB and sequences that contained a pause and final lengthening. To create the latter, five recordings of sequences without an internal IPB were acoustically manipulated on and after the second name. We did not use exactly the same set of recordings without an internal IPB as in Experiments 1 and 2 because some of the sequences contained co-articulation between the final vowel of the second name and the initial vowel of the conjunction such that the insertion of a pause would have created an unnaturally sounding stimulus. Hence, for Experiment 3, we chose five sequences with no or only minimal co-articulation: three sequences that had been used in the previous experiments and two more sequences recorded with the same speaker. First, any co-articulation between the second name and the subsequent conjunction, that is, the section of formant transition from the final vowel [i] to the vowel [u], was cut out at zero crossings. Second, a silent interval of 500 ms – corresponding to the mean duration of pauses measured in natural sequences with an internal IPB from Experiment 1 – was inserted at the offset of the final vowel. Then, the final vowel was lengthened to 180%, according to the average lengthening factor found in the acoustic analysis of sequences with an internal IPB in Experiment 1. A sequence with inserted pause and lengthening cues is depicted in Figure 1D. For both stimulus conditions, the pitch contours were stylized (two semitones) and sequences were resynthesized using the PSOLA function in Praat.

Sequences without an internal IPB lasted on average 1.82 s (range: 1.71–1.89 s), while sequences with inserted pause and lengthening had a mean duration of 2.36 s (range: 2.27–2.42 s). The sound files for the condition without an internal IPB contained seven sequences (two of the five recordings were randomly chosen and repeated at the end of a sound file) and had an average duration of 18.75 s (range: 18.59–18.86 s). To achieve a similar mean duration, the sound files for the condition with inserted pause and lengthening cues contained only six sequences, resulting in an average duration of 19.2 s (range: 19.09–19.24 s). The interstimulus interval between the sequences in each type of sound file was 1 s.

5.3 Procedure

The procedure was identical to that of Experiments 1 and 2. All infants were familiarized with sequences without an internal IPB until at least 20 sequences had been presented. This led to a minimum of 54 s of accumulated listening time. The familiarization was immediately followed by a test phase of twelve test trials, half of them containing familiar sequences without an internal IPB, the other half, containing new sequences with an internal IPB cued by pause and final lengthening.

5.4 Descriptive results

Infants tested in Experiment 3 showed a mean listening time of 8.65 s (SD = 3.97 s) to the novel test trials and a mean listening time of 7.1 s (SD = 3.52 s) to the familiar test trials (see Figure 2). Eleven out of 16 infants had longer listening times to the novel test trials.

6 Joint statistical analysis of the experiments

We statistically analyzed the data of all four experiments in a repeated-measures ANOVA with Familiarity as within-subject factor (mean listening times to novel vs. familiar test trials) and Experiment as between-subject factor (Exp. 1, 2, 2a, 3).[3] This revealed a significant main effect of Familiarity, $F(1,63) = 5.480$, $p = 0.022$, but not of Experiment, $F(3,63) = 1.334$, $p = 0.271$. However, there was a significant interaction of Familiarity and Experiment, $F(3,63) = 5.628$, $p = 0.002$.

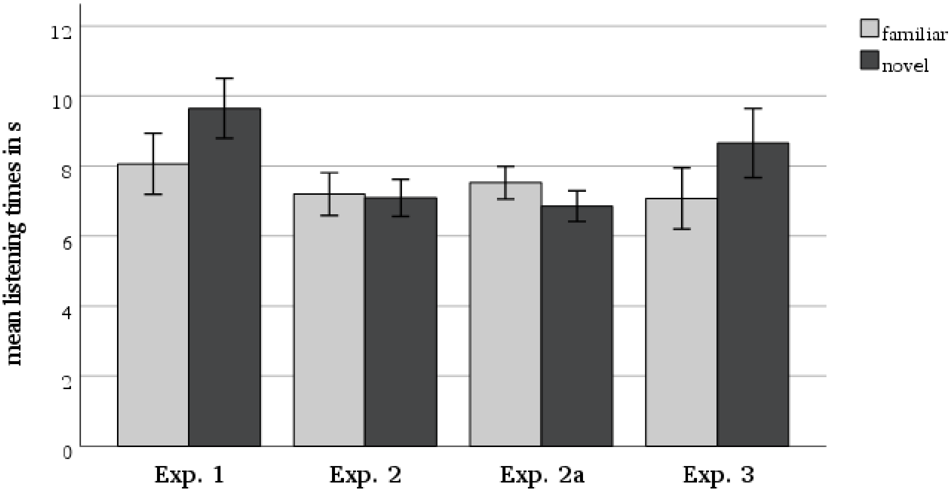


Figure 2: Mean listening times in seconds to familiar and novel test trials after familiarization with sequences without an internal IPB in Experiment 1 (all cues), 2 (pitch and lengthening, short familiarization), 2a (pitch and lengthening, long familiarization) and 3 (pause and lengthening). Error bars indicate ± 1 SE.

To dissolve the significant interaction and to determine which experiments differed from each other we carried out pairwise comparisons. Therefore, we compared the results of each experiment with those from Experiment 2 as a control experiment – the one that yielded the smallest listening time differences between

---

[3] Note that from Experiment 1 only the data from the group familiarized without IPB was considered ($n = 12$).

novel and familiar test trials. We ran a post-hoc t-test on the difference scores (mean listening time to novel test trials minus mean listening time to familiar test trials) in Experiment 2 versus 1, Experiment 2 versus 2a, and Experiment 2 versus 3. Difference scores are depicted in Figure 3.

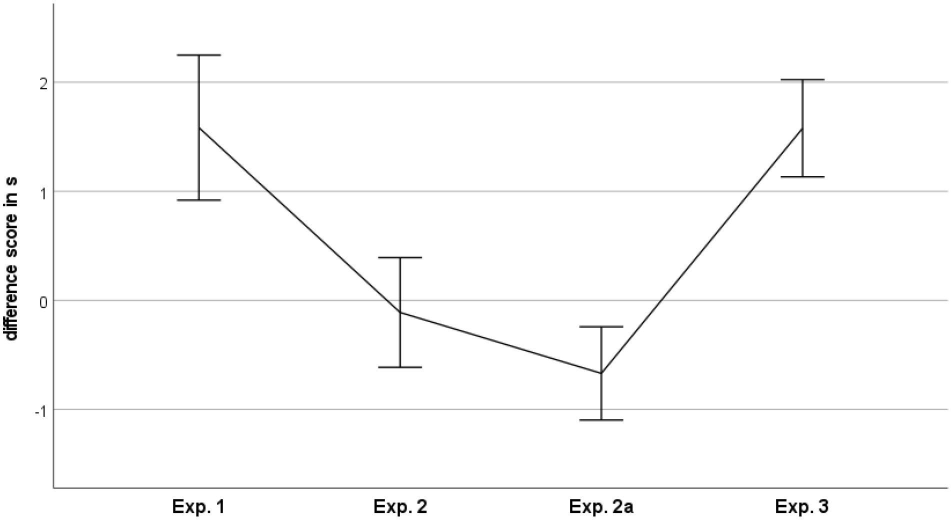


Figure 3: Mean listening time differences to novel minus familiar test trials in seconds. Error bars indicate ± 1 SE.

To adjust for multiple comparisons, the alpha-level of post-hoc t-tests was corrected according to Holm (1979). With three levels of comparisons this resulted in $\alpha = 0.05$ for the largest p-value, $\alpha = 0.025$ for the mid p-value, and $\alpha = 0.017$ for the smallest p-value.

6.1 Pairwise comparison of Experiment 2 versus 1

The post-hoc test for Experiment 2 versus 1 failed to reach significance, M_{Diff} = 1.694 s, $t(26) = -2.073, p = 0.048, \alpha = 0.025$. However, on the descriptive level, we see a much larger listening time difference in Experiment 1 compared to Experiment 2. Infants in Experiment 1 had a mean listening time difference of M_{Diff} = 1.584 s with a preference for novel test trials, whereas infants in Experiment 2 had a mean listening time difference of $M_{Diff} = -0.110$ s. Considering the small sample size that presumably prevents statistical significances, this may indicate that six-month-old infants might tend to discriminate the two types of prosodic patterns when the IPB is indicated by pitch, lengthening, and pause, but not when it is cued by pitch and lengthening only (also see the comparison to Experiment 3 below).

6.2 Pairwise comparison of Experiment 2 versus 2a

The post-hoc test for Experiment 2 versus 2a was not significant, $M_{Diff} = 0.558$ s, $t(37) = 0.843, p = 0.404, \alpha = 0.05$. Infants in Experiment 2a had a mean listening time difference of $M_{Diff} = -0.668$ s with a slight preference for familiar trials. Infants in Experiment 2 had a mean listening time difference of $M_{Diff} = -0.110$ s. This comparison indicates that infants' behavior does not differ between Experiments 2 and 2a. Hence, the data do not support the hypothesis that a longer familiarization phase leads to better discrimination in six-month-olds suggesting that – unlike eight-month-olds – they still need the pause cue to detect the boundary (see Exp. 1). However, it is possible that doubling the familiarization time may have reduced infants' general attention during the test phase and may have obscured their discrimination of the test stimuli. This is also indicated by the high drop-out rate, which suggests the modified version of the experiment was especially hard.

6.3 Pairwise comparison of Experiment 2 versus 3

The post-hoc test for Experiment 2 versus 3 almost reached significance, $M_{Diff} = 1.688$ s, $t(30) = -2.512, p = 0.018, \alpha = 0.017$. Infants in Experiment 3 had a mean listening time difference of $M_{Diff} = 1.578$ s with a preference for novel test trials. Infants in Experiment 2 had a mean listening time difference of $M_{Diff} = -0.110$ s. We interpret this as a tendency towards a better discrimination of stimuli, in which pause and lengthening indicate the boundary, instead of pitch and lengthening. Moreover, the results obtained from Experiment 3 support the interpretation that infants in Experiment 1 detected the boundary that was marked by pause, lengthening, and pitch. Note that the number of participants was higher in Experiment 3 ($n = 16$) compared to Experiment 1 ($n = 12$). This underlines the issue of low statistical power in Experiment 1. Across Experiments 1 and 3, the mean listening time scores were very similar with both revealing a numerically strong novelty effect.

Overall, we interpret the six-month-olds' data as an indicator for successful perception of boundaries that are marked by the full set of cues or by the subset of pause and lengthening. In contrast, the combination of pitch and lengthening seems to be a non-sufficient marking. This points to a crucial role of the pause cue in early prosodic boundary processing in German.

7 General discussion

Perceptual reorganization in early speech perception has been reported extensively and in numerous languages for aspects of segmental phonology. However, research on the development of prosody – specifically phrasal prosody – is still sparse. The present study, in combination with the findings from Wellmann et al. (2012) and Holzgrefe-Lang et al. (2016, 2018), contributes to uncovering a developmental change in the processing of prosodic boundary information in German.

The experiments presented in this paper have addressed the role of pitch change, final lengthening, and pause in boundary detection by German-learning infants. Although the statistical power of our experiments is small due to the low number of infants and the limited amount of trials, our data yield three major results that need to be discussed with caution. First, German six-month-olds (but not four-month-olds) are able to detect a major prosodic boundary signaled by all the cues. Second, pitch change combined with final lengthening did not appear as a sufficient marking for six-month-olds, neither after a prolonged familiarization. Third, six-month-olds do not generally need a combination of all the three cues, but a combination of pause and final lengthening is sufficient to detect the boundary.

We will focus our discussion on the questions that were raised in the introduction: First, what do these results tell us about developmental changes in infant prosodic cue perception? Second, we will embed the results into the previous research on American English- and Dutch-learning infants, focusing on crosslinguistic similarities and differences in prosodic cue weighting. Beyond, we will compare the present behavioral outcomes to electrophysiological findings (Holzgrefe-Lang et al. 2018) and discuss the cognitive demands underlying a potential asymmetry in prosodic cue perception.

7.1 Developmental changes in German boundary perception

The results suggest that German-learning infants have developed a sensitivity to fully marked IPBs by six months[4] and even to a subset of boundary cues containing pause and final lengthening. While six-month-olds heavily rely on the

---

[4] Four-month-old infants displayed a null result in the same Experiment 1 which allows for two interpretations: either, four-month-olds are not yet able to detect fully marked IPBs, or they are able to, but can't show their ability with this kind of method. Even though there are few studies showing that the HPP methods in principle works with four-month-olds (e.g. Bosch & Sebastián-Gallés 2001, Herold et al. 2008, Seidl & Cristià 2008), the familiarization design might not be optimal for this young age group as it depends on a rather high working memory load (process the auditory information and store them to detect the change in the test phase). A study by Höhle et al. (2009) revealed a null result for four-month-olds in a familiarization technique, whereas Herold et al. (2008) evidenced discrimination in the same age group, with the

pause cue to detect a boundary, two months later infants have enhanced their sensitivity to boundaries by perceiving a subtle difference indicated by pitch and lengthening (Wellmann et al. 2012). Perceptual attunement in the acquisition of German phrasal prosody mainly concerns the necessity of the pause cue. Two observations from German, 1) a rather inconsistent occurrence of the pause cue in German adult-directed speech (ADS) as well as 2) no or only short pauses at minor syntactic boundaries, underline the usefulness of a German learner's ability to detect a boundary without pause, and hence support an enhancement in the sensitivity towards boundary cues.

At first glance, the occurrence of a pause seems to be highly reliable with respect to its function as a linguistic boundary cue in infant-directed speech (IDS): whenever pauses occur they are likely to indicate a sentence boundary (Fernald & Simon 1984, Fisher & Tokura 1996). However, a pause does not seem to be the predominant boundary cue in German ADS and occurs only rarely as a single cue: Peters et al. (2005) analysis of phrase boundary markings in the German Kiel Corpus of spontaneous ADS[5] showed that pauses occurred only at 38% of all boundaries, while pitch changes did so at 74% and lengthening at 66%. An essential finding was that cue combination at boundaries was a frequent pattern, occurring at 61.6% of all boundaries. Among these, the co-occurrence of pitch and lengthening (24.6%) and the coalition of all three cues (23.7%) were the most frequent. Cue combinations including only pause and one additional cue were comparatively infrequent: only 8.4% of all boundaries were marked by a combination of pause and lengthening, and 4.9% by a combination of pause and pitch. Each prosodic cue also occurred as a single cue: pitch alone marked 20.8% of all boundaries, while lengthening cued 9.4%, and pause only 1.3%. In brief, Peters et al. (2005) revealed pause to be the least frequent and the combination of pitch and lengthening to be the most frequent marker. This implies that, at least in German ADS, a large proportion of phrase boundaries are not signaled by a pause, which would cause a segmentation problem for learners who overly rely on the occurrence of a pause. Unfortunately, corresponding data on cue frequency in German IDS are missing, but Fernald & Simon (1984) report longer pause duration and a higher correspondence between pause and sentence boundaries in German IDS compared to ADS. Also, a systematic review by Ludusan et al. (2016)

---

same materials, but a change in the experimental setup to a discrimination technique without a familiarization phase. Hence, a more simple preference paradigm might have worked better with our materials in the four-month-olds; however, the data would not be directly comparable to the data of the older infants.

[5] In this analysis, all auditory breaks that occurred turn-internally within the continuous speech stream were classified as phrase boundaries.

across several languages suggests that pause duration is increased in IDS. So, it may be the case that a high reliance on pause as a boundary cue is appropriate when exposed to IDS, but not when exposed to ADS, making a change in the reliance on this cue necessary to become a proficient processor of ADS.

The second argument that relying solely on the pause cue is not an optimal strategy comes from the fact that pauses signal major prosodic and syntactic boundaries, whereas minor prosodic and syntactic boundaries like phrase boundaries are less often marked by a pause (Strangert 1991, Terken & Collier 1992) or they are marked by pauses of shorter durations (e.g., Butcher 1981, Goldman-Eisler 1972). Thus, if prosodic cues are essential for infants for detecting not only major clause boundaries in the signal but also boundaries of smaller units within these larger domains, children must become more sensitive to boundary markings that do not involve a pause. To sum up, the developmental change evidenced here seems to be in line with the requirements of the ambient language German.

7.2 Crosslinguistic comparison

Turning to the next point of the discussion – the crosslinguistic dimension of boundary cue perception – our results reveal similarities as well as dissimilarities in Dutch, German, and American English infants. First, the finding that six-month-old German infants are sensitive to naturally occurring, fully marked IPBs is in line with results from previous studies with Dutch- and American English-learning infants (Johnson & Seidl 2008, Nazzi et al. 2000, Seidl 2007, Soderstrom et al. 2005) and thus expands the crosslinguistic evidence that young infants are sensitive to natural prosodic phrasing. For American English, even four-month-olds have been shown to use fully marked boundaries for the segmentation of complex clauses.

We consider our materials – rather short and phonologically highly-controlled sequences with successively inserted cues – an important extension to the cross-linguistic field of infant prosodic boundary perception. A recent study by van Ommen et al. (2020) created similar stimuli in French to be presented to French- and German-learning infants. The sequences were three coordinated French names either with a major prosodic boundary *[Loulou et Manu][et Nina]* or without a boundary *[Loulou et Manou et Nina]*. Hence, materials were identical in their structure to the concatenation of three German names used in the present study. Also the procedure was the same, a discrimination task in the HPP paradigm with a familiarization phase followed by a test phase. The results showed that French six- and eight-month-olds perceived the boundary when it was signaled by all the three cues, but none of the two age groups was successful when the boundary was cued by pitch and lengthening only. Interestingly, in contrast, German infants presented with the same French materials perceived the pitch-lengthening cued boundary at eight months, but not at six months. This result reinforces the data from the present study and from Wellmann et al. (2012) and supports the interpretation of a developmental change in German boundary perception related to the pause cue. Moreover, no such developmental change can be observed in French infants' performance pointing towards language-specificity of the respective development.

7.2.1 Crosslinguistic similarities

A similarity found across the American English, Dutch, German, and French studies is that pause seems to be a necessary boundary cue for the youngest groups of tested infants (Dutch- and German-learning six- and American English-learning four-month-olds, for French even in both six- and eight-month-olds) – independently of whether the specific task requires a segmentation or a discrimination of stimuli. This points to a language-general way of processing prosodic boundaries in the first months of life that is strongly related to the acoustically salient pause cue.

A strong reliance on the pause is useful since among the three main boundary cues, pause is the most "universal" one. In the languages in which infant boundary perception has been studied, pauses have many pragmatic and paralinguistic functions; however, when it comes to linguistic structure, it serves only one function, that is, the marking of syntactic boundaries. This may render pause a crosslinguistically highly reliable cue. Note that preboundary lengthening and pitch may bear more than one linguistic function. Duration as the acoustic correlate of lengthening is also used to express lexical and/or phrasal stress as well as phonemic contrasts (vowel duration). Regarding pitch, the majority of the world languages are tonal; this means that pitch is used to express different lexical items. In pitch-accent languages like Japanese or stress languages such as English pitch can also be used to distinguish word meanings. Moreover, pitch bears several functions at the sentence level, for example the distinction between declaratives and questions. Moreover, pause is a perceptually rather salient feature of an acoustic signal and it provides categorical information that can be processed locally because the presence or absence of silence can be detected immediately in the signal. This may be different for the other two boundary cues, pitch changes and lengthening, which constitute relational information and require the parsing of longer strings to recognize any changes in pitch and duration at the location of the boundary in relation to the whole speech string.

Therefore, the available results by infants learning American English, Dutch, German, and French revealing that the occurrence of a pause is initially required for boundary detection may reflect a rather universal processing that initially relies on the pause as an acoustically salient categorical cue that can be easily processed independent of the contextual information in every language environment.[6]

7.2.2 Crosslinguistic differences

The major difference in the development between Dutch, American English, German, and French concerns if, and if so, when, infants respond to boundaries that are not marked by a pause. For Dutch, we only know that six-month-olds need the pause for boundary detection since older infants were not tested (Johnson & Seidl 2008). French infants still need the pause cue at eight months (van Ommen et al. 2020). In American English-learning infants, the necessity of a pause as a boundary cue disappears already between the ages of four and six months (Seidl 2007, Seidl & Cristià 2008). This is when in German-learning infants the perception of fully cued boundaries first emerges. Only between the ages of six and eight months a developmental change occurs that makes the pause cue no longer necessary.

Seidl & Cristià (2008) interpret the behavior of the four-month-old American English-learning infants as a so-called holistic processing in which all cues are equally attended to. They argue that this reflects a general processing mechanism rather than a linguistically based strategy. By six months, American English-learning infants do assign more weight to pitch. Seidl and Cristià explain this development through the increased language exposure allowing to observe the distribution of boundary cues in their native language. Infants may have learned by this age, that pauses are unreliable boundary cues, whereas pitch is a more reliable cue to syntactic boundaries in American English.

Comparing the developmental trajectory between American English- and German-learning infants, the data reveal that the development is different at six

---

[6] We are grateful to one reviewer who suggested to link our findings to those of individuals with acquired language impairments in which the special relevance of the pause cue is also evident (Aasland & Baum 2003). When tested on resolving syntactic ambiguities in coordinate structures, a group of individuals with aphasia after left-hemispheric brain damage – in contrast to a control group – was not able to consistently identify the phrase boundary cued by lengthening and pause (with neutral pitch). However, when the pause duration was increased beyond normal ranges, accuracies improved. Thus, pause also seems to play a crucial role in impaired comprehension and may enable boundary detection even in the absence of the pitch cue.

months. American English-learning six-month-olds need the pitch cue (in any combination with another cue), while this seems not to be the case for the German six-month-olds. This may suggest that pitch information is more salient and thus more important for American English learners than for German learners. In fact, crosslinguistic comparisons of pitch in IDS have found that the mean, minimum, and maximum F0 as well as the F0 variability is significantly higher in American English IDS than in German IDS (Fernald & Simon 1984, Fernald et al. 1989). Hence, American English-learning infants may be more prone to attend to pitch variation in their input than German-learning infants.

A comparison of German- and Dutch-learning six-month-olds (Johnson & Seidl 2008) reveals similarities. Like their German age-mates, Dutch-learning infants did not respond to the prosodic boundary marked by pitch change and final lengthening, indicating a crucial role of the pause in Dutch as well. Interestingly, Dutch – like German – IDS was found to show a lower mean and a lower range in F0 difference compared to American English IDS (Fernald et al. 1989, van de Weijer 1997), suggesting again that the properties of the specific speech input relate to crosslinguistic differences in how infants process prosodic information and that with less pitch variation pauses may become a more crucial cue for the marking and the perception of prosodic boundaries.

At eight months, German infants' sensitivity has developed to perceiving a pitch-lengthening cued boundary to such an extent that it can even be applied in a non-native language (van Ommen et al. 2020). The result that French eight-month-olds' boundary detection still depends on all three cues points to a delay in comparison to the German-learning infants. This is supported by van Ommen et al.'s experiment with French and German adults who did not differ in their discrimination of sequences with pitch and lengthening only. Apparently, French listeners catch up at one point. van Ommen et al. (2020) argue that the language-specific differences at eight months might stem from a higher prosodic variability in German providing a larger basis to attend to prosodic details. French does not use prosodic characteristics to mark lexical stress. It uses prosody for phrasal stress; however, phrasal stress coincides with phrasal boundaries by default. Hereby, French is highly regular in the employment of prosodic cues. German, on the contrary, employs a larger variety of tonal and duration patterns at the phrasal as well as at the lexical level, and these are not strictly aligned to boundaries. This might explain why the German-learning infants show an earlier sensitivity to the specific cue combinations than their French peers.

The comparisons between the German, American English, Dutch, and French studies (that, notably, varied in the experimental paradigms: discrimination vs. segmentation) only give first, still vague indications of crosslinguistic differences in prosodic boundary cue weighting that support the assumption that perceptual attunement occurs in this domain. Future research using more comparable materials and methods across languages is necessary to provide a reliable picture of potential crosslinguistic effects of boundary perception and their development. In addition, it is not clear whether the few studies so far have used acoustic instantiations of the different cues that are typical for the specific language and typical for the infants' input. Corresponding prosodic analyses of ADS and IDS in the respective languages are therefore needed to broaden our understanding of the early prosodic development. A further limitation in the interpretation of the results concerns the strength of single cues, which might differ between different cue constellations. There is evidence that marking of prosodic boundaries is subject to cue trading relations, that is, an interaction between the strength of the cues that mark the boundary with one cue being stronger when another cue is weaker (e.g., Beach 1991). The cue insertion applied to the German stimuli was based on the acoustic parameters of pitch, lengthening, and pause that had been measured in natural sequences with a fully marked IPB. Although these values were already quite high (a pitch rise of 212 Hz/13.65 semitones, a final lengthening factor of 1.8, and a pause of 500 ms duration; cf. Peters (2005)[7]), they might be even higher in natural sequences with a boundary that is only marked by the subset of pitch and lengthening; in other words, when pause is missing, the other cues may be enhanced. Thus, we cannot exclude that six-month-olds might also be able to detect a boundary without pause if we had implemented stronger pitch and/or lengthening cues. Given that continuous stimulus manipulations can hardly be investigated in behavioral tasks in infants, the present experimental design did not consider cue trading relations. Still, we can conclude that developmental changes in behavior occur, since the eight-month-olds were able to detect the boundary using identical materials.

7.3 Behavioral versus neurophysiological methods

In the final section, we compare the present behavioral finding to those of a previous electrophysiological study. Using the very same stimuli, Holzgrefe-Lang et al. (2018) investigated boundary perception in eight- and also six-month-old German-learning infants by means of event-related potentials (ERP). In adults, the processing of IPBs with and without a pause evokes a specific ERP component, the so-called closure positive shift (CPS; e.g., Holzgrefe-Lang et al. 2016,

---

[7] In a perception study with adult listeners, Peters (2005) implemented lengthening factors of 1.2, 1.4, and 1.6 and pauses with a duration continuum between 50 ms and 890 ms – based on the values found in German ADS (Peters et al. 2005)

Steinhauer et al. 1999), which is assumed to reflect the perception of a prosodic boundary. Holzgrefe-Lang et al. (2018) investigated whether an infant CPS can be elicited in response to different cue constellations. Specifically, they compared six- and eight-month-old infants' brain response to stimuli containing either no boundary cue, a combination of pitch change and lengthening, or only a pitch cue. The ERPs in response to the latter condition did not differ from the condition without any boundary cues, but the combined occurrence of pitch change and final lengthening elicited a positivity that resembled the adult CPS in both age groups (Holzgrefe-Lang et al. 2016). Hence, the electrophysiological data suggests that six- and eight-month-old German infants do not differ in IPB perception, whereas the current HPP data provides no evidence that six-month-olds detect pitch and lengthening cued boundaries, but suggests a developmental change between the ages of six and eight months. Thus, prosodic boundary perception without the pause cue is evidenced earlier at the electrophysiological level (but see Männel & Friederici 2009, Männel et al. 2013 for data that indicate that stimuli with neutralized pause cues would only elicit a CPS in children older than three years). In line with this asymmetry, there is ample evidence from other studies (Friederici et al. 2007, Höhle et al. 2009, Schipke 2012) that a specific brain response may precede the corresponding behavioral response in the course of development. For instance, the recognition of the ambient language's dominant stress pattern has been shown for four-month-old German learners using ERPs (Friederici et al. 2007), whereas a behavioral preference is evident only at six months (Höhle et al. 2009).

We assume that the diverging results across the different methods are due to different cognitive demands during testing. ERPs are measured on-line during infants' passive listening, and hence do not depend on task demands or an overt response performance involving additional processing requirements (see Männel & Friederici 2008). In the case of the ERP study by Holzgrefe-Lang et al. (2018), the brain response indicating the perception of a prosodic boundary marked by pitch and lengthening occurs right after the presentation of the phrase-final syllable. Hence, the ERPs represent immediate processing responses evoked by the presence of specific boundary cues in the stimuli. In the HPP, the infants' behavioral response is measured by the amount of listening time indicated by the infant's head turn towards the side of presentation. The expectation to observe differences in listening times between the conditions in the HPP experiment is based on the assumption that a representation of the familiarization stimulus has been formed during the familiarization phase and that the stimuli presented during the test phase are mapped onto this representation. Establishing this representation requires that at least some memory traces survive the switch to the test phase of the experiment. Such a long-lasting memory component is not involved in the ERP paradigm. Nevertheless, the fact that the six-month-olds show effects in the HPP when the boundary is marked by the full set of boundary cues or by the subset of pause and lengthening suggests that memory requirements alone are not sufficient to explain the null effect in the condition with only lengthening and pitch. Rather, differences in the level of attention might account for the different outcomes across the ERP and HPP measurements as well as for the different outcomes across ages in the behavioral studies. Considering that listening times are an indicator of attention, the change that we observed between the six- and the eight-month-olds in the HPP data may suggest that infants have sharpened their attention towards pitch and lengthening, which are functionally relevant cues to German phrase boundaries, at the age of eight months.

8 Conclusion

To conclude, the present study provides evidence that German six-month-old infants are able to detect a major prosodic boundary characterized by the three main cues. This ability is crucial for the first steps in language acquisition as it equips the naïve learner with a tool to chunk the continuous speech stream into clauses. In a headturn preference procedure discrimination task, we found that for six-month-olds pitch and lengthening cues are not sufficient, but they need the pause cue. Boundary detection on the basis of combined relational prosodic cues like pitch changes and final lengthening shows up only by eight months. We argued that this behavioral change displays an enhancement in sensitivity that is reflected in a shift of attention to boundary markings that are functionally relevant in the ambient language. The ability to detect a boundary with the full as well as with a subset of cues enables syntactic parsing of not only major, but also minor, syntactic units. This ability is also necessary for the adult listener, especially in the case of structural ambiguities. Therefore, being able to detect these boundaries with their language-specific markers is essential to becoming an efficient processor of a given language.

Acknowledgements

We thank T. Fritzsche for technical assistance in the Potsdam BabyLab. Thanks to A. Beyer, S. Fischer, B. Graf, T. Leitner, M. Orschinsky, and M. Zielina for their help in recruiting and testing the infants. We are grateful to R. Räling for stimuli production. Thanks to all parents and their children who participated in this study. Many thanks to S. Jähnigen and B. Smolibocki for help with the acoustic manipulation of the stimuli.

Funding information

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), priority program SPP 1234, project no. 18181517 (GZ HO 1960/13-1 and FR2865/2-1).

References

Aasland, Wendi A. & Shari R. Baum. 2003. Temporal parameters as cues to phrasal boundaries: A comparison of processing by left- and right-hemisphere brain-damaged individuals. *Brain and Language* 87(3). 385–399. DOI: 10.1016/S0093-934X(03)00138-X.

Beach, Cheryl M. 1991. The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language* 30(6). 644–663. DOI: 10.1016/0749-596X(91)90030-N.

Best, Catherine C. & Gerald W. McRoberts. 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech* 46(Pt 2-3). 183–216. DOI: 10.1177/00238309030460020701.

Bhatara, Anjali, Natalie Boll-Avetisyan, Annika Unger, Thierry Nazzi & Barbara Höhle. 2013. Native language affects rhythmic grouping of speech. *The Journal of the Acoustical Society of America* 134(5). 3828–43. DOI: 10.1121/1.4823848.

Bijeljac-Babic, Ranka, Josette Serres, Barbara Höhle & Thierry Nazzi. 2012. Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PloS ONE* 7(2). 1–8. DOI: 10.1371/journal.pone.0030843.

Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer*. http://www.praat.org/.

Bosch, Laura & Núria Sebastián-Gallés. 2001. Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy* 2(1). 29–49. DOI: 10.1207/S15327078IN0201_3.

Butcher, Andy. 1981. *Aspects of the speech pause: Phonetic correlates and communicative functions*. Universität Kiel. (Doctoral dissertation).

Fernald, Anne & Thomas Simon. 1984. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology* 20(1). 104–113. DOI: 10.1037/0012-1649.20.1.104.

Fernald, Anne, Traute Taeschner, Judy Dunn, Mechthild Papousek, Bénédicte de Boysson-Bardies & Ikuko Fukui. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16(3). 477–501. DOI: 10.1017/s0305000900010679.

Féry, Caroline & Gerrit Kentner. 2010. The prosody of embedded coordinations in German and Hindi. In *Proceedings of Speech Prosody*, 1–4. https://www.isca-speech.org/archive/pdfs/speechprosody_2010/fery10_speechprosody.pdf.

Fisher, Cynthia & Hisayo Tokura. 1996. Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development* 67(6). 3192–3218. DOI: 10.2307/1131774.

Friederici, Angela D., Manuela Friedrich & Anne Christophe. 2007. Brain responses in 4-month-old infants are already language specific. *Current Biology* 17(14). 1208–11. DOI: 10.1016/j.cub.2007.06.011.

Goldman-Eisler, Frieda. 1972. Pauses, clauses, sentences. *Language and Speech* 15(2). 103–113. DOI: 10.1177/002383097201500201.

Götz, Antonia, H. Henny Yeung, Anna Krasotkina, Gudrun Schwarzer & Barbara Höhle. 2018. Perceptual reorganization of lexical tones: Effects of age and experimental procedure. *Frontiers in Psychology* 9. DOI: 10.3389/fpsyg.2018.00477.

Gout, Ariel, Anne Christophe & James L. Morgan. 2004. Phonological phrase boundaries constrain lexical access ii. infant data. *Journal of Memory and Language* 51(4). 548–567. DOI: 10.1016/j.jml.2004.07.002.

Herold, Birgit, Barbara Höhle, Elisabeth Walch, Tanja Weber & Michael Obladen. 2008. Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life. *Developmental Medicine and Child Neurology* 50(9). 678–83. DOI: 10.1111/j.1469-8749.2008.03055.x.

Hirsh-Pasek, Kathy, Deborah G. Kemler Nelson, Peter W. Jusczyk, Kimberly Wright Cassidy, Benjamin Druss & Lori Kennedy. 1987. Clauses are perceptual units for young infants. *Cognition* 26(3). 269–286. DOI: 10.1016/S0010-0277(87)80002-1.

Hirst, Daniel & Albert Di Cristo. 1998. A survey of intonation systems. In Daniel Hirst & Albert Di Cristo (eds.), *Intonation systems: A survey of twenty languages*, chap. 1, 1–44. Cambridge: Cambridge University Press.

Höhle, Barbara, Ranka Bijeljac-Babic, Birgit Herold, Jürgen Weissenborn & Thierry Nazzi. 2009. Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development* 32(3). 262–274. DOI: 10.1016/j.infbeh.2009.03.004.

Holm, Sture. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6. 65–70. http://www.jstor.org/stable/4615733.

Holzgrefe-Lang, Julia, Caroline Wellmann, Barbara Höhle & Isabell Wartenburger. 2018. Infants' processing of prosodic cues: Electrophysiological evidence for boundary perception beyond pause detection. *Language and Speech* 61(1). 153–169. DOI: 10.1177/0023830917730590.

Holzgrefe-Lang, Julia, Caroline Wellmann, Caterina Petrone, Romy Räling, Hubert Truckenbrodt, Barbara Höhle & Isabell Wartenburger. 2016. How pitch change and final lengthening cue boundary perception in German: Converging evidence from ERPs and prosodic judgements. *Language, Cognition and Neuroscience* 31(7). 904–920. DOI: 10.1080/23273798.2016.1157195.

Hunter, Michael A. & Elinor W. Ames. 1988. A multifactoral model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research* 5. 69–95.

Huttenlauch, Clara, Marie Hansen, Carola de Beer, Sandra Hanne & Isabell Wartenburger. 2023. Age effects on linguistic prosody in coordinates produced to varying interlocutors: Comparison of younger and older speakers. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 157–192. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777534.

Johnson, Elizabeth K. & Amanda Seidl. 2008. Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy* 13(5). 440–455. DOI: 10.1080/15250000802329321.

Jusczyk, Peter W., Anne Cutler & Nancy J. Redanz. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development* 64(3). 675–87. DOI: 10.1111/j.1467-8624.1993.tb02935.x.

Kemler Nelson, Deborah G., Kathy Hirsh-Pasek, Peter W. Jusczyk & Kimberly Wright Cassidy. 1989. How the prosodic cues in motherese might assist language learning. *Journal of Child Language* 16(1). 55–68. DOI: 10.1017/s030500090001343x.

Kohler, Klaus J. 1983. Prosodic boundary signals in German. *Phonetica* 40(2). 445–472. DOI: 10.1159/000261685.

Kuhl, Patricia K., Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani & Paul Iverson. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9(2). F13–F21. DOI: 10.1111/j.1467-7687.2006.00468.x.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Bjorn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255. 606–608. DOI: 10.1126/science.1736364.

Lehiste, Ilse, Joseph P. Olive & Lynn A. Streeter. 1976. Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America* 60(5). 1199–1202. DOI: 10.1121/1.381180.

Ludusan, Bogdan, Alejandrina Cristia, Andrew Martin, Reiko Mazuka & Emmanuel Dupoux. 2016. Learnability of prosodic boundaries: Is infant-directed speech easier? *The Journal of the Acoustical Society of America* 140(2). 1239–1250. DOI: 10.1121/1.4960576.

Männel, Claudia & Angela D. Friederici. 2008. Event-related brain potentials as a window to children's language processing. In Irina A. Sekerina, Eva M. Fernandez & Harald Clahsen (eds.), *Developmental psycholinguistics. On-line methods in children's language processing*, 29–72. Amsterdam: Benjamins. DOI: 10.1075/lald.44.04man.

Männel, Claudia & Angela D. Friederici. 2009. Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. *Journal of Cognitive Neuroscience* 21(10). 1988–2006. DOI: 10.1162/jocn.2009.21221.

Männel, Claudia, Christine S. Schipke & Angela D. Friederici. 2013. The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience* 5. 86–94. DOI: 10.1016/j.dcn.2013.01.003.

Mattock, Karen & Denis Burnham. 2006. Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy* 10(3). 241–265. DOI: 10.1207/s15327078in1003_3.

Mattock, Karen, Monika Molnar, Linda Polka & Denis Burnham. 2008. The developmental course of lexical tone perception in the first year of life. *Cognition* 106(3). 1367–1381. DOI: 10.1016/j.cognition.2007.07.002.

Maurer, Daphne & Janet F. Werker. 2014. Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology* 56(2). 154–178. DOI: 10.1002/dev.21177.

Nazzi, Thierry, Deborah G. Kemler Nelson, Peter W. Jusczyk & Ann Marie Jusczyk. 2000. Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy* 1(1). 123–147. DOI: 10.1207/S15327078IN0101_11.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris Publications. DOI: 10.1515/9783110977790.

Peters, Benno. 2005. Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache. In Klaus J. Kohler, Felicitas Kleber & Benno Peters (eds.), *Prosodic structures in German spontaneous speech*, 203–345. Kiel: Arbeitsberichte (AIPUK). https://www.ipds.uni-kiel.de/kjk/pub_exx/aipuk35a/aipuk35a_8.pdf.

Peters, Benno, Klaus J. Kohler & Thomas Wesener. 2005. Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In Klaus J. Kohler, Felicitas Kleber & Benno Peters (eds.), *Prosodic structures in German spontaneous* *speech*, 143–184. Kiel: Arbeitsberichte (AIPUK). https://www.ipds.uni-kiel.de/kjk/pub_exx/aipuk35a/aipuk35a_6.pdf.

Petrone, Caterina, Hubert Truckenbrodt, Caroline Wellmann, Julia Holzgrefe-Lang, Isabell Wartenburger & Barbara Höhle. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61. 71–92. DOI: 10.1016/j.wocn.2017.01.002.

Polka, Linda & Janet F. Werker. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20(2). 421–435. DOI: 10.1037//0096-1523.20.2.421.

Price, Patti J., Mari Ostendorf, Stefanie Shattuck-Hufnagel & Cynthia Fong. 1991. The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America* 90(6). 2956–70. DOI: 10.1121/1.401770.

Sanderman, Angelien A. & René Collier. 1997. Prosodic phrasing and comprehension. *Language and Speech* 40(4). 391–409. DOI: 10.1177/002383099704000405.

Schipke, Christine S. 2012. *Processing mechanisms of argument structure and case-marking in child development: Neural correlates and behavioral evidence*. Max Planck Institute for Human Cognitive & Brain Sciences. (Doctoral dissertation). http://pubman.mpdl.mpg.de/pubman/item/escidoc:1481652/component/escidoc:1760308/139_Schipke.pdf.

Schmitz, Michaela. 2008. *The perception of clauses in 6- and 8-month-old German-learning infants: Influence of pause duration and the natural pause hierarchy*. University of Potsdam. (Doctoral dissertation). http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-29078.

Schubö, Fabian & Sabine Zerbian. 2023. The patterns of pre-boundary lengthening in German. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 1–34. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777526.

Scott, Donia R. 1982. Duration as a cue to the perception of a phrase boundary. *The Journal of the Acoustical Society of America* 71(4). 996–1007. DOI: 10.1121/1.387581.

Seidl, Amanda. 2007. Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language* 57(1). 24–48. DOI: 10.1016/j.jml.2006.10.004.

Seidl, Amanda & Alejandrina Cristià. 2008. Developmental changes in the weighting of prosodic cues. *Developmental Science* 11(4). 596–606. DOI: 10.1111/j.1467-7687.2008.00704.x.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology* 3(1986). 371–405. DOI: 10.1017/S0952675700000695.

Selkirk, Elisabeth. 2005. Comments on intonational phrasing in English. In Sonia Frota, Marina Vigário & Maria Joao Freitas (eds.), *Prosodies: With special reference to Iberian languages*, 11–58. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110197587.

Shukla, Mohinish, Katherine S. White & Richard N. Aslin. 2011. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences USA* 108(15). 6038–43. DOI: 10.1073/pnas.1017617108.

Skoruppa, Katrin, Ferran Pons, Anne Christophe, Laura Bosch, Emmanuel Dupoux, Núria Sebastián-Gallés, Rita Alves Limissuri & Sharon Peperkamp. 2009. Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science* 12(6). 914–919. DOI: 10.1111/j.1467-7687.2009.00835.x.

Soderstrom, Melanie, Deborah G. Kemler Nelson & Peter W. Jusczyk. 2005. Six-month-olds recognize clauses embedded in different passages of fluent speech. *Infant Behavior and Development* 28(1). 87–94. DOI: 10.1016/j.infbeh.2004.07.001.

Steinhauer, Karsten, Kai Alter & Angela D. Friederici. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience* 2(2). 191–196. DOI: 10.1038/5757.

Strangert, Eva. 1991. Where do pauses occur in text read aloud? In Halldór Ármann Sigurðsson (ed.), *Papers from the Twelfth Scandinavian Conference of Linguistics*, 403–414. Reykjavik: Linguistic Institute, University of Iceland.

Streeter, Lynn A. 1978. Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America* 64(6). 1582–1592. DOI: 10.1121/1.382142.

Terken, Jacques & René Collier. 1992. Syntactic influences on prosody. In Yoh'ichi Tokhura, Eric Vatikiotis-Bateson & Yoshinori Sagisaki (eds.), *Speech perception, production and linguistic structure*, 427–438. Amsterdam: IOS Press.

Truckenbrodt, Hubert. 2005. A short report on intonation phrase boundaries in German. *Linguistische Berichte* 203. 273–296.

Truckenbrodt, Hubert. 2007a. The syntax-phonology interface. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 435–456. Cambridge: Cambridge University Press.

Truckenbrodt, Hubert. 2007b. Upstep on edge tones and on nuclear accents. In Carlos Gussenhoven & Tomas Riad (eds.), *Tones and tunes: Experimental studies in word and sentence prosody*, 349–386. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110207576.2.349.

Truckenbrodt, Hubert. 2016. Focus, intonation, and tonal height. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, 463–482. Oxford: Oxford University Press.

Vaissière, Jacqueline. 1983. Language-independent prosodic features. In Anne Cutler & D. Robert Ladd (eds.), *Prosody: Models and measurements*, 53–66. Berlin: Springer. DOI: 10.1007/978-3-642-69103-4_5.

van de Weijer, Joost. 1997. Language input to a prelingual infant. In *Proceedings of the GALA '97 conference on language acquisition*, 290–293. Edinburgh: Edinburgh University Press.

van Ommen, Sandrien, Natalie Boll-Avetisyan, Saioa Larraza, Caroline Wellmann, Ranka Bijeljac-Babic, Barbara Höhle & Thierry Nazzi. 2020. Language-specific prosodic acquisition: A comparison of phrase boundary perception by French- and German-learning infants. *Journal of Memory and Language* 112. DOI: 10.1016/j.jml.2020.104108.

Wellmann, Caroline, Julia Holzgrefe, Hubert Truckenbrodt, Isabell Wartenburger & Barbara Höhle. 2012. How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology* 3. 580. DOI: 10.3389/fpsyg.2012.00580.

Werker, Janet F. & Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7(1). 49–63. DOI: 10.1016/S0163-6383(84)80022-3.

Yeung, H. Henny, Ke Heng Chen & Janet F. Werker. 2013. When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language* 68(2). DOI: 10.1016/j.jml.2012.09.004.

Zhang, Xinting. 2012. *A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese*. University of Michigan. (Doctoral dissertation). https://hdl.handle.net/2027.42/96107.

Chapter 5

Age effects on linguistic prosody in coordinates produced to varying interlocutors: Comparison of younger and older speakers

Clara Huttenlauch[a], Marie Hansen[a], Carola de Beer[a,b], Sandra Hanne[a] & Isabell Wartenburger[a]

[a]University of Potsdam [b]University of Bielefeld

This production study builds on and extends the research on how prosodic cues can be used to resolve syntactic ambiguities. We compared how younger speakers (mean age 25 years, Huttenlauch et al. 2021) and older speakers (mean age 68 years) produced prosodic cues to distinguish between structurally different coordinated three-name sequences without and with internal grouping of the first two names. The prosodic cues of interest were variations in f0 (F0 range), duration of segments at the end of the names (final lengthening), and pause insertion. In line with the Proximity/Similarity model by Kentner & Féry (2013), we found that both age groups used all three cues to signal the grouping: Prosodic cues were modified on the group-internal Name1 as well as on Name2 at the right-most element of the group. These prosodic cues were clearly understood by naïve listeners. Successful prosodic disambiguation was not affected by age-related differences in speech production. Furthermore, we analysed the productions with regard to different contexts, such as addressing interlocutors of different ages and mother tongues, and in noisy environments. We found that both age groups of speakers used the same prosodic cues consistently across all contexts, indicating that the use of prosodic cues to clarify syntactic ambiguities is a stable part of the production process, which we interpret as being in line with models of situational independence of disambiguating prosody (e.g., Schafer et al. 2000). Our study provides evidence that the use of these prosodic cues (F0 range, final lengthening, and pause) is a reliable way to clarify ambiguous structures in speech and independent of the speaker's age.

Clara Huttenlauch, Marie Hansen, Carola de Beer, Sandra Hanne & Isabell Wartenburger. 2023. Age effects on linguistic prosody in coordinates produced to varying interlocutors: Comparison of younger and older speakers. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 157–192. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777534

1 Introduction

Linguistic prosody, as in prosodic boundaries, can be used to resolve syntactic ambiguities. Such syntactic ambiguities exist in coordinated sequences of more than two elements (e.g., names) since those elements can be grouped internally at different levels. For instance, the three-name sequence *Moni and Lilli and Manu* can describe three individual persons or a group of three persons (i.e., no internal grouping as in (1)) or a group of two persons in addition to one individual person, with two different possibilities for the grouping (i.e., the group can consist of *Moni and Lilli* or of *Lilli and Manu.* (2) gives an example for the internal grouping of *Moni and Lilli* indicated by parentheses). The latter two different groupings correspond to underlying syntactic structures that differ in their direction of embedding. The difference to the first sequence is the depth of embedding. The absence or type of internal grouping as in (1) versus (2) in an answer to the question *Who will plant a tree?* results in either one, or two, or three planted trees. Prosody, thus, brings the underlying structure to the surface (i.e., disambiguates the otherwise ambiguous surface structure). In this study, we will compare productions of a structure without internal grouping (1) to a structure with internal grouping of the first two elements (2).

(1) Name1 and Name2 and Name3. – without internal grouping

(2) (Name1 and Name2) and Name3. – with internal grouping

1.1 Prosodic marking in coordinate sequences

In German, the difference between the two structures (i.e., the resolution of the structural ambiguity) is mainly indicated by one or more of three prosodic cues: F0 change, final lengthening, and pause (Peters et al. 2005, Gollrad et al. 2010, Kentner & Féry 2013, Petrone et al. 2017, for final lengthening see also Schubö & Zerbian 2023 [this volume]). Young speakers have been shown to use these three prosodic cues to clearly mark the internal grouping of coordinated name sequences (Kentner & Féry 2013, Petrone et al. 2017, Huttenlauch et al. 2021). Figure 1 provides visualisations of waveform and spectrogram with F0 contour and segmental annotations of productions without and with internal grouping, respectively, generated using Praat (Boersma & Weenink 2019). The marking of the internal grouping appears as a global and not a local phenomenon, in accordance with the Proximity/Similarity model (Kentner & Féry 2013): Young speakers modified prosodic cues not only at the right edge of the internal group (i.e., on

Name2 in the example in (2)), but already earlier in the utterance (i.e., on Name1, see also left and right panel in Figure 1, Kentner & Féry 2013, Huttenlauch et al. 2021). The principle of Proximity relates to the syntactic constituent structure (Kentner & Féry 2013). The proximity of syntactically grouped elements is expressed by a weakening of the prosodic cues (e.g., less final lengthening, lower F0 peak, smaller F0 range) on the left-most element of two sister elements (e.g., Name1 in (2), Moni in right panel of Figure 1) compared to an ungrouped element in the same position (e.g., Name1 in (1), Moni in left panel of Figure 1). The principle of Anti-Proximity predicts a strengthening of the prosodic cues (e.g., more final lengthening, higher F0 peak, larger F0 range, insertion of a pause) on/after the right-most element of a group than on/after an ungrouped element (e.g., Name2 in (2) versus in (1), Lilli in right versus left panel of Figure 1). The principle of Similarity relates to the depth of syntactic embedding and since it does not apply to our structures we will not discuss it further. In summary, in name sequences with grouping such as (2), the productions of Name1 contain weaker prosodic cues and those of Name2 encompass stronger prosodic cues compared to name sequences without grouping such as (1).

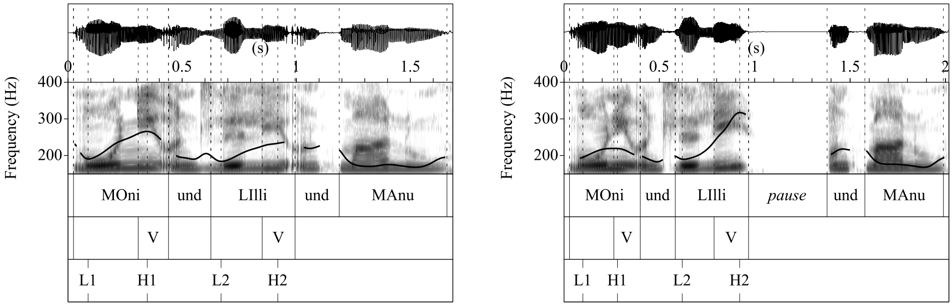


Figure 1: Waveform and spectrogram with F0 contour (black line) of the coordinated name sequence *MOni und LIlli und MAnu* (capital letters correspond to stressed syllable) produced without internal grouping (left) and with internal grouping (right) by a young female speaker. The TextGrid gives an example for the manual annotation of low (L) and high (H) F0 values and the segmentation of the final vowels within Name1 and Name2.

In perception, the early cues on Name1 could reliably be recovered to predict the upcoming structure by more than half of the participants in a two-alternative forced choice decision task with gated stimuli (Hansen et al. 2022). Although all young speakers in Huttenlauch et al. (2021) reliably marked the constituent grouping of coordinated names, they showed inter-speaker variability in how they phonetically realised the prosodic boundary, especially final lengthening was used in a more flexible way than F0 range and pause. Besides prosodic disambiguation, Huttenlauch et al. (2021) investigated the situational (in)dependence of disambiguating prosody by comparing prosodic cues addressed to interlocutors differing in age and mother tongue as well as in the absence/presence of background noise. Despite the phonetic variability in the realisation of prosodic cues between speakers, the data show a rather consistent pattern of prosodic cues across different communicative situations. The latter finding was interpreted as indexing situational independence: Disambiguating prosody seems to be produced automatically by the speakers in a rather invariant manner.

The present study builds on and extends the results on prosodic boundary production of young speakers (Huttenlauch et al. 2021) with productions of older speakers. Data of both age groups were elicited with the same design and materials, which allows for a direct comparison and detailed investigation of age effects. Age has not only been shown to affect language production in terms of word-finding abilities (for a review see Burke & Shafto 2004) but also in terms of altered acoustic characteristics affecting prosody-related features in the tonal and durational domain. Age, thus, has an effect on the same features that are relevant for the realisation of linguistic prosody.[1] Age, therefore, may interact with the modulation of prosodic cues in conveying the intended meaning. In the remaining part of the introduction, we will address age-related changes in the tonal and durational domain in general (Section 1.2) and their possible impact on the use of linguistic prosody in particular (Section 1.3). Finally, we will present findings on the situational (in)dependence of prosodic cues (Section 1.4).

1.2 Age-related changes in the tonal and durational domain in general

In the following section, we will summarise previous research on general age-related changes in the tonal and durational domains. It is important to note that studies differ in how they group participants into age ranges and in how many years each age group spans. We will use *young* or *younger speakers* to refer to the age range between 18 and 30 years of age and *older speakers* for ages above 60 years.

In the tonal domain, age effects on fundamental frequency (F0) have been studied for several measures including mean and median F0, the span between minimum and maximum (F0 range), and the variability of those measures captured in standard deviations (SD). Here, we focus on the latter two as mean or median F0 are rather uninformative in the context of our study, which focuses on

---

[1] We are aware of the multitude of non-linguistic information transmitted through prosodic cues including but not limited to the emotional state and background of the speaker. In the context of this study, we are only interested in linguistic prosody.

analysing F0 range. So far, results are inconclusive and in part divergent between genders. For F0 range, some studies report no differences between younger and older speakers (Markó & Bóna 2010, Smiljanic & Gilbert 2017), while Dimitrova et al. (2018), Tuomainen & Hazan (2018), and Hazan et al. (2019) observed a larger F0 range for older compared to younger women and Kemper et al. (1998) found a smaller F0 range in older compared to younger speakers irrespective of gender. When it comes to F0 variability, there is evidence for an increase with increasing age (Scukanec et al. 1992, Lortie et al. 2015, Santos et al. 2021). More variability and less stability in older speakers compared to younger speakers was further noticed by several studies looking at more specific measures regarding speech acoustics (including jitter, shimmer, and noise-to-harmonics-ratio; Goy et al. 2013, Lortie et al. 2015, Rojas et al. 2020 among others).

In the durational domain, previous studies observed slower speaking/articulation rates in older compared to younger speakers (Tuomainen & Hazan 2018, Hazan et al. 2019, Tuomainen et al. 2019, 2021 and references in a review by Tucker et al. 2021: 5), relating this finding mainly to longer syllable or word durations (Scukanec et al. 1996, Harnsberger et al. 2008, Barnes 2013, Dimitrova et al. 2018), longer segment durations (Kemper et al. 1995, Harnsberger et al. 2008, Smiljanic & Gilbert 2017), or an increased number of pauses (Kemper et al. 1998, Dimitrova et al. 2018). However, no evidence for pause duration as a driver of age-related differences in speech rate has been reported so far (Barnes 2013, Smiljanic & Gilbert 2017, Dimitrova et al. 2018).

To sum up, previous researchers provided some evidence for tonal and durational differences between younger and older speakers, indicating increased F0 ranges and durations with increasing age. Since these changes affect the same channel used to convey linguistic meaning, we will address possible interferences in the next paragraph.

1.3 Age-related changes in the tonal and durational domain alongside linguistic prosody

We will now turn towards studies that can help to address the question of whether age-related changes in the tonal and durational domain interact with the modulation of disambiguating prosodic cues, as these studies used speech material that explicitly required the use of linguistic prosody. Scukanec et al. (1996) measured the maximal F0 value within the vowel of elicited monosyllabic words in either contrastive or non-contrastive stress position in younger and older female English speakers. Both age groups used F0 in a similar way to mark the focused words (Scukanec et al. 1996: 235). However, independent of the word position in the sentence, older speakers produced higher F0 values than young speakers in words with contrastive stress and lower maximal F0 values in words in non-contrast positions. The authors concluded that, for the analysed data set, age did not influence the productions of "linguistically salient variations in prosodic output" (Scukanec et al. 1996: 238). The difference in the maximal F0 values between words with and without contrastive stress was even larger in older than in young speakers. The same holds true for the durational domain: Even though older speakers produced longer word durations together with larger standard deviations (i.e., more variability), both age groups used duration to linguistically distinguish stressed from unstressed words.

Further evidence that older speakers use lengthening for prosodic disambiguation despite an overall age-related slower speaking rate comes from Tauber et al. (2010) and Barnes (2013) who reported longer durations for older English speakers in disambiguating contexts. Barnes (2013) elicited structurally ambiguous sentences with either high or low attachment of the prepositional phrase (e.g., *The girl hit the boy with the fan*) in younger and older English speakers. Although the study found longer durations of the direct object and the prepositional phrase regardless of target in the productions of older speakers than in the productions of younger speakers, the overall results revealed that both age groups used the prosodic cues mean F0, pause duration, word duration, and mean intensity similarly to disambiguate ambiguous sentences. However, in another task tapping production of lexical stress to differentiate noun-verb pairs with strong-weak and weak-strong stress patterns, "older adults utilised F0 to a significantly greater extent than young adults" (Barnes 2013: 43). Tauber and colleagues elicited structurally ambiguous sentences (e.g., *The lake froze over a month ago*) to explicitly test for age differences in the realisation of disambiguating prosody in English sentences (Tauber et al. 2010). They found that intonational boundaries (defined as pause duration plus duration of the critical word at the boundary) were longer in older than in younger speakers. Notably, both age groups seem to have had difficulties with the task, as the percentage of sentences which were successfully disambiguated via prosody was 66% for older speakers (above chance, $p < 0.05$) and 59% for the younger age group (not significantly above chance, $p > 0.06$) (Tauber et al. 2010).

In summary, even though age leads to changes in the tonal and temporal domain in general, there is evidence from English speakers that the modulation of prosody to convey linguistic meaning remains unaffected. Older participants even appear to produce prosodic cues in a more extreme way than younger speakers. To the best of our knowledge, there is no study that addressed age differences in the use of prosody to resolve ambiguities in coordinate structures. If the findings for English ambiguous sentences are transferable to German coordinate structures, we expect that older speakers disambiguate coordinate structures using more extreme prosodic cues than young speakers. This motivates our first research question:

RQ1: Prosodic disambiguation of coordinate name sequences: Do older speakers compared to young speakers show a more extreme use of the three prosodic cues F0 range, final lengthening, and pause on Name1 and Name2 to mark the internal grouping of coordinates in German?

1.4 Situational (in)dependence of prosodic cues

In the remaining part of the introduction, we will address the situational (in)dependence of prosodic cues, a second topic investigated in Huttenlauch et al. (2021). It deals with the effects of different types of interlocutors and the absence/presence of noise on the use of disambiguating prosodic cues. Huttenlauch et al. (2021) compared the use of prosodic cues in five *contexts* involving four female interlocutors: a young adult (YOUNG), a child (CHILD), an elderly adult (ELDERLY), and a young non-native speaker of German (NON-NATIVE) and in noise (the young adult with background white noise, NOISE). The productions directed at the young adult native speaker (i.e., the context YOUNG) were taken as a baseline for comparisons. The findings showed stability in the use of prosodic cues for disambiguating the internal structure of coordinates. That is, individual speakers produced a limited set of cue patterns with only slight shifts in cue distribution across different contexts. This stability in prosodic patterns for disambiguation irrespective of the context was interpreted in favour of models of situational independence of disambiguating prosody (Schafer et al. 2000, Kraljic & Brennan 2005, Speer et al. 2011). These models predict that disambiguating prosody is produced in an automatic way, for the sake of the speakers themselves, and hence depends neither on the presence or absence of an interlocutor, nor on the type of interlocutor or situational setting (e.g., background noise). Despite arguing for situational independence of disambiguating prosody, Huttenlauch et al. (2021) found slight prosodic modifications in the data that can be attributed to context effects. Similarly, as discussed for the prosodic marking of internal grouping of coordinates in the first part of the introduction, the question arises whether age effects in the tonal and durational domain have an impact on the use of F0 range, final lengthening, and pause when speaking in different contexts and whether we find age effects in the situational (in)dependence of prosodic disambiguation. Research on age effects in speech production to different interlocutors is, to our knowledge, still scarce. In the following, we will briefly summarise existing findings including the context effects found in the productions of young speakers in Huttenlauch et al. (2021).

With regard to addressing a child interlocutor, we will refrain from summarising the immense body of literature treating speech towards preverbal infants since the use of prosody for disambiguation requires that language ability has already been acquired to a certain extent. We are not aware of studies investigating effects of speaker age on prosodic cues uttered towards a child interlocutor. For young speakers, speech towards a child interlocutor has been described as containing an increased F0 range (Biersack et al. 2005, Huttenlauch et al. 2021), lengthened vowels (Biersack et al. 2005), or more pauses (DePaulo & Coleman 1986).

Speech addressing an elderly interlocutor has been explored in data on young and older adult speakers. While younger speakers slowed down their speaking rate by increasing vowel duration and inserting more pauses in speech addressing an elderly interlocutor, older speakers did not do so (Kemper et al. 1995). For older speakers addressing a young interlocutor, however, Kemper and colleagues observed a slower speaking rate than for young speakers. The authors argued that, in comparison to young speakers, older speakers adopt a more simplified speech style including lower speaking rate when addressing a young interlocutor, and thus it is possibly hard for them to slow down even further in order to adapt to an elderly interlocutor (Kemper et al. 1995: 56). Furthermore, young speakers addressing an elderly interlocutor, slowed down their speaking rate with longer pauses, increased final lengthening (Huttenlauch et al. 2021), and increased F0 range or variation in F0 (Thimm et al. 1998, Huttenlauch et al. 2021).

We are not aware of studies investigating effects of speaker age on prosodic cues when addressing a non-native interlocutor. Some studies involving young speakers found no clear differences (DePaulo & Coleman 1986, Uther et al. 2007, Knoll & Scharrer 2007, Knoll et al. 2011, Huttenlauch et al. 2021), while others observed a lowered speech rate due to lengthened pauses (Biersack et al. 2005), a higher mean F0 (Knoll et al. 2015), increased word durations and intensity (Rodriguez-Cuadrado et al. 2018), or an increased F0 range along with segmental modifications described as a more emphatic style (Smith 2007; see Piazza et al. 2021 for a review on foreigner-directed speech).

Finally, speech in noisy environments compared to silent environments is affected by modulations in several ways. The reported changes are referred to as "Lombard speech" (Lombard 1911 as cited in Zollinger & Brumm 2011) and include decreased speaking rate (due to increased segment or word durations), increased

F0 ranges, increased signal amplitude, and spectral changes such as smaller spectral slope (e.g., Junqua 1996, Summers et al. 1988, Jessen et al. 2003, Zollinger & Brumm 2011, Smiljanic & Gilbert 2017, Tuomainen et al. 2019, 2021). The findings for young speakers in a noisy environment in Huttenlauch et al. (2021) were interpreted as being partly in line with Lombard speech, as they revealed increased final lengthening and decreased pause duration but no changes in F0 range. With respect to age effects in speech adaptation to noise, no age differences were found by Dromey & Scott (2016) and Smiljanic & Gilbert (2017), with the latter reporting an age-independent decrease in speaking rate when noise was present, while Tuomainen et al. (2019) reported a decreased speaking rate only for the older age group.

To summarise, the modifications of prosodic cues in coordinates induced by varying contexts observed by Huttenlauch et al. (2021) were rather small but in line with previous findings. The effect of age on the realisation of prosodic cues in more communicative settings with varying interlocutors is still only scarcely explored. For the reported age-related changes in addressing different interlocutors, the question arises whether they replicate to coordinate structures in German. Given the limited evidence, we keep our second research question rather open:

RQ2: Situational (in)dependence: Do young and older speakers differ in adapting their use of prosodic cues when addressing varying interlocutors?

In the current study, we extend the age range of usually studied participants (in Huttenlauch et al. 2021 19–34 years) to older people aged between 60 and 80 years of age (i.e., comparable to the older age groups in the previously presented literature) and compare the productions of linguistic prosody in young and older adult speakers. Specifically, we explore whether age interacts with the modulation of prosodic cues, especially F0 range, final lengthening, and pause, and whether any such interaction may impact the disambiguation of structurally ambiguous coordinated name sequences and the use of prosodic cues when addressing different interlocutors (i.e., regarding situational (in)dependence of disambiguating prosody).

2 Methods and material

Methods, materials, and data of the younger speakers are taken from Huttenlauch et al. (2021) and extended by the data of older speakers.

2.1 Participants

Fifteen young monolingual German native speakers (13 female, 1 male, 1 other; age range: 19–34, mean 25.47 years, SD: 4.6; see Huttenlauch et al. 2021) and 13 older monolingual German native speakers (9 female, 3 male, 1 no information; age range: 61–80 years, mean: 67.77 years, SD: 6.8) were included in the study. Additional five speakers took part in the study, but were discarded due to low task compliance ($n = 1$), scores below 25 in the Montreal Cognitive Assessment (Nasreddine et al. 2005) ($n = 3$), or missing data ($n = 1$). All participants (henceforth *speakers*) were recruited in Potsdam, Germany, and were reimbursed or received course credits (the latter only applies to the young speakers). They were naïve to the purpose of the study and gave written consent to participate. The Ethics Committee of the University of Potsdam approved the procedure of this study (approval number 72/2016). Hearing ability was assessed by a hearing screening using an audiometer (Hortmann DA 324 series) and calculated following the grades of hearing impairment by the WHO as reported in Olusanya et al. (2019). Normal hearing was defined as an average pure-tone audiometry of 25 dB HL or better of 500, 1000, 2000, and 4000 Hz in the better ear. Following this definition, all 15 young speakers and 10 of the older speakers had normal hearing, the remaining speakers showed a slight ($n = 2$) or moderate impairment ($n = 1$).

2.2 Stimuli

2.2.1 Items

As stimuli, we used the same six coordinated name sequences as in Holzgrefe-Lang et al. (2016), Huttenlauch et al. (2021), and Wellmann et al. 2023 [this volume]: Each sequence consisted of three German names coordinated by *und* (English 'and') that appeared in each of two conditions: without internal grouping (3) or with internal grouping of the first two names (4). The grouping of the first two names was visually indicated to the participants by bracketing Name1 and Name2 with parentheses as in (4). The conditions will henceforth be referred to as *brack* for the condition with internal grouping and *nobrack* for the condition without internal grouping. A total of 12 items was used. Young speakers produced each item once per context (see Section 2.2.2), older speakers twice to enlarge the data set and to increase statistical power.

(3) Name1 and Name2 and Name3. Moni und Lilli und Manu.

(4) (Name1 and Name2) and Name3. (Moni und Lilli) und Manu.

The set of coordinates comprised nine different German names in total, all of which were controlled for number of syllables (disyllabic), stress pattern (penultimate), and sonority of the segments (only sonorant material, to facilitate pitch tracking). Six of the names featured the high frontal vowel /i/ in word-final position (Moni, Lilli, Leni, Nelli, Mimmi, and Manni) in order to decrease glottalisation and occurred as Name1 or as Name2. Name3 contained either /u/ or /a/ in word-final position (Manu, Nina, and Lola). Regarding possible collocations of the selected names for each coordinate, there was no particular co-occurrence of two adjacent names (as in, e.g., "Bonnie and Clyde") in the dlexDB corpora (Heister et al. 2011) or in printed sources between 1500 and 2021, as ascertained by the Google Ngram Viewer (Lin et al. 2012).

2.2.2 Contexts

Five different communicative contexts (YOUNG, CHILD, ELDERLY, NON-NATIVE, NOISE) were created that differed in the interlocutor and/or the absence/presence of background white noise (see Table 1). Speakers saw their interlocutors on a screen in two short videos each (one with a personal introduction of the interlocutor and one with instructions for the task) to get an audio-visual impression. The young and non-native interlocutors were similar in age to the group of young speakers, the elderly interlocutor was two years older than the oldest speaker in the group of older speakers. A more detailed description of the videos and interlocutors can be found in Huttenlauch et al. (2021).

2.3 Procedure

Productions were elicited by means of a referential communication task. Contexts were presented blockwise, always starting with the YOUNG context, which served as a baseline in the analysis. The order of the other four contexts was randomised. Each block started with the two video clips of the corresponding interlocutor. Then, for each trial, speakers first saw a fixation cross on the screen accompanied with the auditory presentation of the trigger question *Wer kommt?* ('Who is coming?') via headphones produced by the interlocutor of the current block as a reminder to whom they were talking. After 1000 ms, the fixation cross was replaced by the visual presentation of the name sequence (i.e., the item) in one of the two conditions (see Figure 2). The task was to produce the item in a way that would allow the interlocutor "to understand as rapidly and accurately as possible who is coming together". Recordings took place in a sound-attenuated booth at the University of Potsdam via an Alesis iO/2 audio interface using an AKG HSC271 headset with over-ear headphones and a condenser microphone.

Table 1: Fictional names, ages, origins, and further information of the interlocutors present in the five contexts.

	YOUNG (baseline)	CHILD	ELDERLY	NON-NATIVE	NOISE
Name:	Hannah	Carlotta	Maria Korbmacher	Zsófi	Hannah + white noise
Age (in years):	24	6	82	26	See YOUNG
Origin:	Eberswalde	Potsdam	NA	NA	
Residence:	Potsdam	Potsdam	Potsdam	Potsdam	
Occupation:	Biology student	School child	Retired school teacher	Exchange student	
Further facts:	Moved to Potsdam for her studies, lives in a shared flat, likes the parks in Potsdam	Likes horse riding, her parents pick her up from school, is good at swimming	Lives for two years in an old-age home with her husband, tends to forget things from time to time	Started to learn German one year ago, lives in a shared flat, enjoys doing sports	

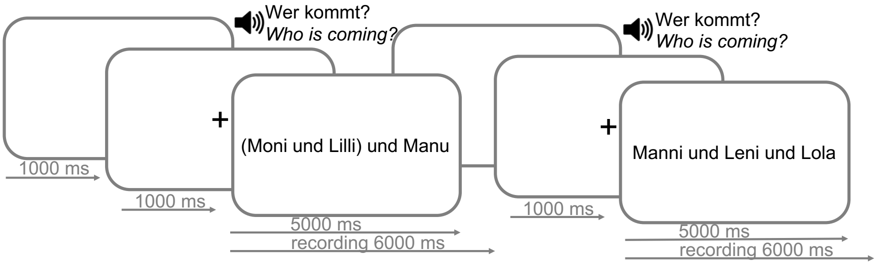


Figure 2: Experimental setting and timing of two trials.

The wide screen in the recordings booth had a resolution of 1920×1200, stimuli were written in Arial, font size 50. The experiment was run from a Dell laptop using Presentation software (*Neurobehavioural Systems* 2018). Each item was presented in each context once (for young speakers) or twice (for older speakers). Thus, the data set contained 900 individual productions of young speakers (6 name sequences × 2 conditions × 5 contexts × 15 young speakers) and 1560 individual productions of older speakers (6 name sequences × 2 conditions × 5 contexts × 2 repetitions × 13 speakers).

2.4 Perception check

After data collection of the production study, all recordings were auditorily presented to naïve listeners who were asked to indicate for each production the perceived condition. To this end they were given two pictograms with three persons each, one pictogram per condition (Figure 3, picture A without and picture B with internal grouping).

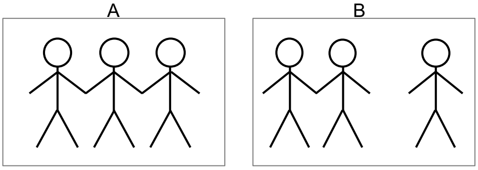


Figure 3: Pictograms used in the perception check depicting the condition without grouping (left panel) and with grouping (right panel).

The aim of the perception check was to assess whether naïve listeners perceive the grouping of the coordinates in the way it was *intended*. By *intended* we refer to the indication of condition which was given to speakers by parentheses around the grouped names in the production study. Obviously, the intention of speakers at the time of the production remains unknown to us.

The data of the young and older age groups were rated separately. The recordings were distributed across different lists with 147 to 267 items. Each listener judged one list and each list was judged by seven or eight listeners.

The perception check of the productions of the young speakers was conducted in presence of several listeners in the same room with a paper-and-pen version. Data of 31 listeners (22 female, 9 male; age range: 18–41, mean: 24.1 years, SD: 5.8) were analysed. Another 11 listeners took part in the study, but had to be excluded due to technical problems ($n = 9$), German as a non-native language ($n = 1$) or a hit-rate 2 SD below the mean hit-rate of all listeners ($n = 1$, see Huttenlauch et al. 2021 for more details).

For the productions of the older speakers, the perception check was transferred onto OpenSesame (Mathôt et al. 2012) and was run as a web-based study on JATOS (Lange et al. 2015) in individual sessions. Data of 49 listeners (29 female, 9 male, 11 other/no information; age range: 18–63, mean: 24.63 years, SD: 6.3) were analysed. Another five listeners took part in the study, but had to be excluded due to technical problems.

In the analysis of the perception check, the exclusion threshold for individual productions was set to a hit-ratio 2 SD below the mean ratio, as suggested by standard assumptions on the exclusion of data points (e.g., Howell et al. 1998). Hit-ratio was calculated separately for each production as the number of congruent rates (i.e., correct identification of the intended grouping/condition, referred to as *hit-rate*) divided by the number of total rates. Applying this criterion, 36 productions (4%, 11 nobrack, 25 brack) in the group of the young speakers and 66 productions (4%, 39 nobrack, 27 brack) in the group of the older speakers fell below the threshold and were excluded from further analyses. For a more detailed description of procedure and analysis of the perception check see Huttenlauch et al. (2021).

2.5 Segmentation and measurements

In addition to the productions excluded based on the perception check, three productions were excluded from analysis in the data set of the older speakers: due to hesitations that made the analysis of condition impossible ($n = 2$) and due to recording problems ($n = 1$). The final data set comprised 2355 productions (young: 864, older: 1491). Table 2 provides an overview of how the productions distribute across age groups, conditions, and contexts.

Table 2: Distribution of productions entering statistical analyses across age groups, conditions, and contexts in the final data set.

age group	condition	YOUNG	CHILD	ELDERLY	NON-NATIVE	NOISE
younger	nobrack	87	85	90	90	87
	brack	83	88	89	86	79
older	nobrack	141	148	148	151	153
	brack	151	153	153	148	145

For the extraction of the three prosodic cues under investigation, segment boundaries and pauses were manually annotated in Praat (Boersma & Weenink 2019, version 6.0.32) by following the criteria in Turk et al. (2006). Silent intervals of at least 20 ms duration were considered as pauses (following the procedure in Petrone et al. 2017). F0-minima (L) and F0-maxima (H) on both Name1 and Name2, were manually annotated (example TextGrids are given in Figure 1). The points were set into parts of the signal, where F0 can be reliably measured (i.e., avoiding the edges of segments, glottalised parts in the signal, and parts with other non-modal voice quality). The F0 contour mostly displayed a rising movement on Name1 and Name2, respectively (i.e., L preceded H). Only in a few cases, speakers produced a falling F0 movement on Name1 (young speakers: 88 falls versus 776 rises, older speakers: 108 falls versus 1368 rises) or Name2 (older speakers: 13 falls versus 1458 rises). For some productions in the data of the elderly speakers it was impossible to find reliable locations to annotate either L and/or H points and it was, thus, impossible to measure the F0 range. In those cases, the corresponding item was excluded from the analysis of F0 range for Name1 and/or Name2. This applies to 15 items (1.0% of the productions of older speakers) in the condition without internal grouping and to 20 items (1.3% of the productions of older speakers) with internal grouping. All in all, we aimed for an approach of measuring F0 range that was applicable to the majority of the recordings. For further segmentation criteria see Huttenlauch et al. (2021). For Name1 and Name2 separately, we calculated the three variables F0 range, final lengthening, and pause. The variable F0 range reflects the range between the F0-minimum and the F0-maximum on NameX in semitones (st; calculated as $12 \times \log_2(F0_H/F0_L)$). The variable final lengthening reflects the duration of the final vowel of NameX divided by the duration of NameX (in %, the final vowel is annotated as *V* on the second tier of the TextGrid in Figure 1.). The pause variable reflects the duration of a possible pause after NameX divided by the duration of the whole utterance (in %). We chose relative instead of absolute measures as they are independent of individual speech rates and mean fundamental frequency. However, to descriptively assess potential age-related effects, absolute durational measurements were taken into consideration.

2.6 Statistical analysis

The workflow of the statistical analyses was similar to that in Huttenlauch et al. (2021), additionally comprising a group comparison between young and older speakers. For each dependent variable (F0 range, final lengthening, pause) on Name1 and Name2, we ran separate linear mixed-effects regression models in R (R Core Team 2018). Each model estimated the difference in the dependent variables between the two age groups (young and older speakers), between the four context comparisons, and between the two conditions (brack and nobrack), if applicable. Interactions between context and age group were added to further explore the dependencies of the differences, as well as interactions of context and age group with condition. A maximal model including all main effects and their interactions, as previously described, as well as including a random effects structure with all possible variance components and correlation parameters associated with the four within-subject contrasts (child vs. young, elderly vs. young, non-native vs. young, noise vs. young) was always fit first.[2] In order to avoid overfitting of the random effects structure, we followed the approach outlined in Bates et al. (2015) and conducted an iterative reduction of model complexity. A more detailed explanation of the model reduction, along with all reduced models and the complete model outputs of the fixed effects, can be found on an Open Science Framework project page (https://osf.io/fc8nz) together with the data and code. In the results section, we will only report the statistically significant effects which comprise main effects of condition and/or main effects and interactions of age group.

3 Results

In the following, we will first present descriptive results from absolute and relative measurements with a focus on age, including a statistical comparison of the age groups. Hereafter, we will turn towards the results of linear mixed models fit to compare the age groups regarding their use of prosodic cues for disambiguation (RQ1) and regarding their adaptation to different interlocutors (RQ2).

3.1 Descriptive statistics and statistical age group comparison of absolute durational measurements

In the main section of our analysis, we analysed the use of prosodic cues by measuring the relative duration of speech segments and pauses. This method allowed us to understand how prosodic cues were used, regardless of individual

---

[2]
```
Prosodic_cue ~ 1 + condition*context*age_group +
      (1 + condition +
      child_vs_young + elderly_vs_young + nonnat_vs_young + noise_vs_young +
      age_group +
      condition:age_group +
      condition:child_vs_young + condition:elderly_vs_young +
      condition:nonnative_vs_young + condition:noise_vs_young +
      child_vs_young:age_group + elderly_vs_young:age_group +
      nonnative_vs_young:age_group + noise_vs_young:age_group +
      condition:child_vs_young:age_group + condition:elderly_vs_young:age_group +
      condition:nonnative_vs_young:age_group +
      condition:noise_vs_young:age_group | speaker)
```

differences in speaking rate or the absolute duration of sounds. Before presenting the relative measurements, we will present some absolute durational measurements to compare the differences between younger and older speakers (cf. Table 3). However, we will not include measurements of average F0 by age group because the speaker groups had mixed genders, which could affect our estimation of differences in F0 between the groups.

Table 3: Descriptive statistics of absolute durational measurements by age group and statistical group comparison.

	Younger		Older		Comparison
Measurement (ms)	mean	SD	mean	SD	*p*
utterance duration	1964.63	292.16	2181.25	444.80	< 0.0001
final vowel duration (Name1)	129.61	40.09	144.68	46.20	< 0.0001
pause duration (after Name2)	172.93	195.24	262.83	330.05	< 0.0001
final vowel duration (Name2)	181.53	59.51	198.24	65.57	< 0.0001

In our data set we observe longer absolute durations for older as compared to younger speakers for the whole utterance (mean difference of 217 ms), the final vowels of Name1 and Name2 (mean difference of 15 ms and 17 ms, respectively), and the pause after Name2 (mean difference of 89.9 ms). All age group comparisons were statistically significant in linear models with age group as a single sum-contrasted predictor (0.5 for young and −0.5 for older speakers). Moreover, we observe a higher degree of variation (larger SDs) for older speakers than for young speakers across all durational measurements.

3.2 Descriptive statistics of relative measurements

Relative measurements of F0 range, final lengthening, and pause were used to explore the use of prosodic cues for the disambiguation of coordinates with and without internal grouping. Figure 4 shows a visual description of mean location and spread of F0 range as well as final lengthening on Name1 by age group, context, and condition. For both cues and for each context, the mean values in the brack condition are lower for younger than for older speakers, while in the nobrack condition in all contexts except YOUNG, the mean values are larger for younger compared to older speakers. Considering these raw data visually, the difference between conditions is larger in the productions of young speakers than in that of older speakers. We did not run statistical analyses and do not

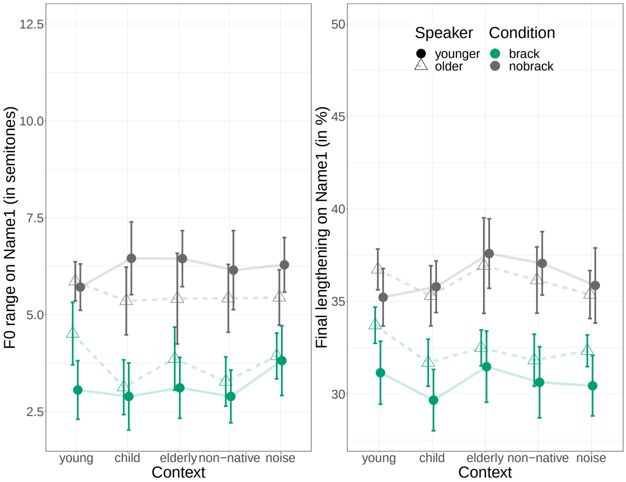


Figure 4: Distribution of raw values of F0 range (left panel) and final lengthening (right panel) on Name1 (y-axis) divided by context (x-axis), condition (colour: grey for nobrack, green for brack), and age group (shape: circles for young speakers, triangles for older speakers). Whiskers show 95% confidence intervals.

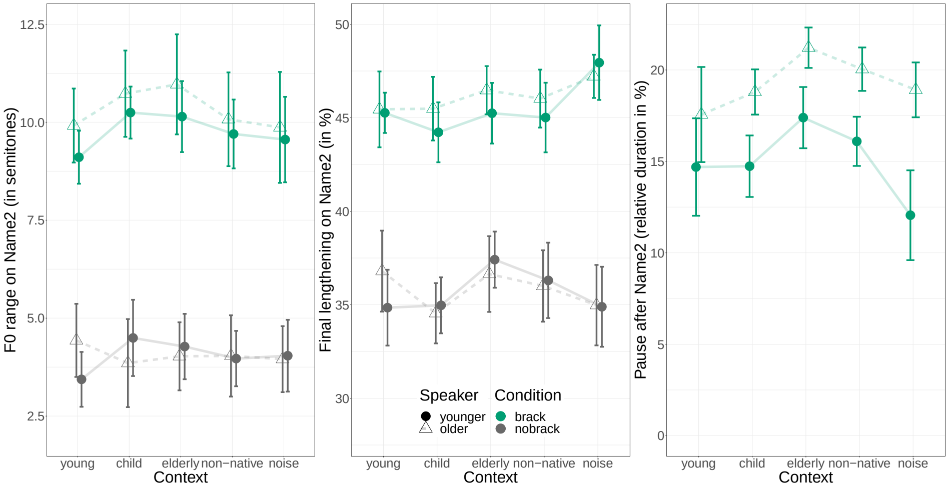


Figure 5: Distribution of raw values of F0 range (left panel), final lengthening (mid panel), and pause (right panel) on Name2 (y-axis) divided by context (x-axis), condition (colour: grey for nobrack, green for brack), and age group (shape: circles for young speakers, triangles for older speakers). Whiskers show 95% confidence intervals.

report descriptive statistics on pause duration after Name1 since mostly zero values were produced by the participants. That is, a pause after Name1 was only produced in 206 out of 2355 trials in total, 175 times in the nobrack condition and 31 times in the brack condition. Figure 5 shows a visual description of mean location and spread of F0 range, final lengthening, and pause on/after Name2 by age group, context, and condition. There is no apparent visual pattern that would apply to both speaker groups and all three cues. For F0 range and pause in the brack condition, young speakers produced smaller mean values than older speakers. For final lengthening in general and F0 range of the nobrack condition, the values are more mixed between age groups. With regard to the direction of the difference in the degree of F0 range and final lengthening between the brack and nobrack condition, both prosodic cues show smaller values in brack than in nobrack on Name1 and the opposite pattern, larger values in brack than in nobrack, on Name2.

To summarise, a visual inspection of the raw data reveals differences between the two age groups in the amount to which the different prosodic cues were produced in the respective contexts and conditions. Nevertheless, the general patterns for each cue are quite similar across contexts for both, young and older speakers. That is, for instance for F0 range in the brack condition in Figure 5 (left panel, green data points), the connecting lines between contexts have slopes in the same directions between speaker groups and in any case do not cross. We are aware that the descriptive analysis of the data does not allow for any generalisations. In the following sections, we will present the results of the statistical models we ran on each cue and Name individually.

3.3 Statistical analyses on Name1

3.3.1 F0 range on Name1

Results for F0 range on Name1 are reported from a reduced model[3] (all final models and code can be found on https://osf.io/fc8nz). Several effects were statistically significant (see Table 4 and https://osf.io/fc8nz).

---

[3]
```
F0_name1 ~ 1 + condition*context*age_group +
          (1 + child_vs_young + elderly_vs_young + noise_vs_young +
           age_group +
           condition:age_group +
           nonnative_vs_young:age_group +
           condition:child_vs_young:age_group +
           condition:nonnative_vs_young:age_group | speaker)
```

Table 4: Selected model estimates and 95% confidence intervals of the fixed effects for F0 range on Name1 including main effect of condition and main effect and interactions of age group. * $p < 0.05$; ** $p < 0.01$.

Predictor	Estimate	95% CI
Intercept	4.666^{**}	(4.060, 5.273)
condition	-1.236^{**}	(−1.559, −0.913)
age group	0.002	(−1.211, 1.216)
condition:age group	−0.593	(−1.239, 0.053)
child vs. young:age group	1.225^{**}	(0.563, 1.886)
elderly vs. young:age group	0.757	(−0.259, 1.773)
non-native vs. young:age group	0.928^{*}	(0.217, 1.639)
noise vs. young:age group	1.193^{*}	(0.230, 2.155)
condition:child vs. young:age group	0.051	(−0.515, 0.616)
condition:elderly vs. young:age group	−0.013	(−0.450, 0.423)
condition:non-native vs. young:age group	0.130	(−0.404, 0.664)
condition:noise vs. young:age group	0.271	(−0.170, 0.712)

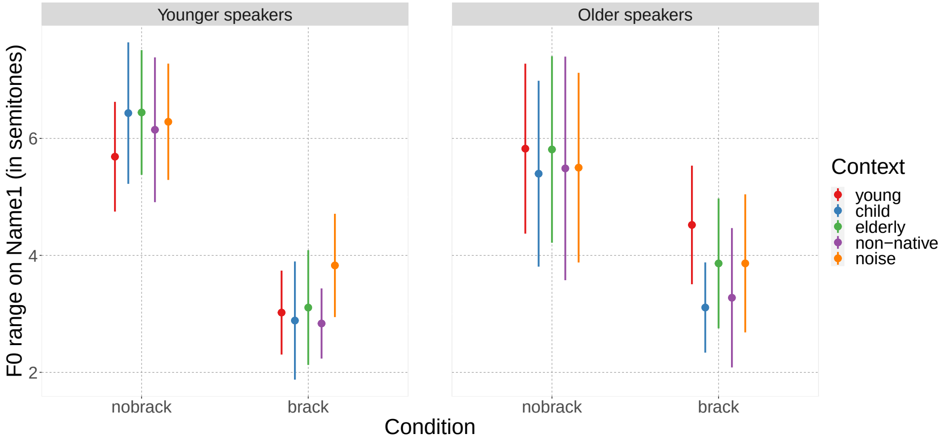


Figure 6: Model predictions for F0 range on Name1 (y-axis) divided by age group (younger speakers left panel, older speakers right panel), condition (x-axis), and context (colour). Whiskers show 95% confidence intervals.

The statistically significant main effect of condition ($\beta = -1.236, p < 0.0001$) confirms that F0 range was used for the disambiguation of brack and nobrack on Name1 by speakers of both age groups: The F0 range in the brack condition was decreased by about 2.5 semitones compared to the nobrack condition. With respect to age-related differences in situational (in)dependence, the statistically significant two-way interactions of the context comparisons CHILD vs. YOUNG ($\beta = 1.225, p = 0.0003$), NON-NATIVE vs. YOUNG ($\beta = 0.928, p = 0.011$), and NOISE vs. YOUNG ($\beta = 1.193, p = 0.016$) with age group, respectively, indicate general age-related differences when addressing the child and non-native as compared to the young interlocutor, as well as age-related differences in noisy vs. non-noisy settings with a young interlocutor. In all of the three context comparisons, young speakers increased their F0 range compared to context YOUNG, while older speakers decreased their F0 range. Model predictions for F0 range on Name1 by condition, context, and age group are displayed in Figure 6.

3.3.2 Final lengthening on Name1

Results for final lengthening on Name1 are reported from a reduced model.[4] Several effects were statistically significant (see Table 5 and link in Section 2.6). The statistically significant main effect of condition ($\beta = -2.366, p < 0.0001$) confirms that final lengthening was used for the disambiguation of brack and nobrack on Name1 by speakers of both age groups: Final lengthening was decreased in the brack condition (where the final vowel span about 31% of the total name duration) as compared to the nobrack condition (where the final vowel span about 36% of the total name duration). With respect to age-related differences in situational (in)dependence, the statistically significant two-way interaction of the context comparison CHILD vs. YOUNG with age group ($\beta = 1.449, p = 0.002$) indicates that young speakers, in contrast to older speakers, increased final lengthening when addressing the child compared to the young interlocutor. A similar pattern is predicted by the model for the context comparison NON-NATIVE vs. YOUNG, for which the interaction with age group was statistically significant ($\beta = 1.877$, $p = 0.028$): While final lengthening is increased by young speakers when addressing the non-native as compared to the young interlocutor, final lengthening is decreased by older speakers. Model predictions for final lengthening on Name1 by condition, context, and age group are displayed in Figure 7.

---

[4] The model can be found at https://osf.io/fc8nz.

Table 5: Selected model estimates and 95% confidence intervals of the fixed effects for final lengthening on Name1 including main effect of condition and main effect and interactions of age group. $^{*}\ p < 0.05$; $^{**}p < 0.01$.

Predictor	Estimate	95% CI
Intercept	33.848**	(32.716, 34.980)
condition	−2.366**	(−2.949, −1.784)
age group	−0.794	(−3.058, 1.469)
condition:age group	−1.001	(−2.166, 0.164)
CHILD VS. YOUNG:age group	1.449*	(0.063, 2.834)
ELDERLY VS. YOUNG:age group	1.962	(−0.255, 4.179)
NON-NATIVE VS. YOUNG:age group	1.877*	(0.203, 3.551)
NOISE VS. YOUNG:age group	1.371	(−0.289, 3.032)
condition:CHILD VS. YOUNG:age group	−0.841	(−2.226, 0.545)
condition:ELDERLY VS. YOUNG:age group	−0.361	(−1.740, 1.017)
condition:NON-NATIVE VS. YOUNG:age group	−0.612	(−1.995, 0.771)
condition:NOISE VS. YOUNG:age group	−0.726	(−2.124, 0.672)

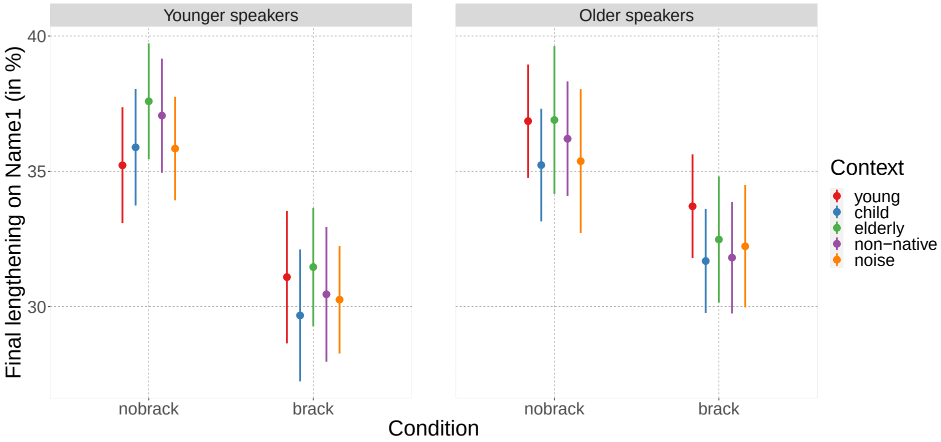


Figure 7: Model predictions for final lengthening on Name1 (y-axis) divided by age group (younger speakers left panel, older speakers right panel), condition (x-axis), and context (colour). Whiskers show 95% confidence intervals.

3.4 Statistical analyses on Name2

3.4.1 F0 range on Name2

Results for F0 range on Name2 are reported from a reduced model.[5] Several effects were statistically significant (see Table 6 and link in Section 2.6). The statistically significant main effect of condition ($\beta = 3.04, p < 0.0001$) confirms that F0 range was used for the disambiguation of brack and nobrack on Name2 across both age groups: The F0 range in the brack condition was increased by about six semitones compared to the nobrack condition. With respect to age-related differences in situational (in)dependence, the significant two-way interaction of the context comparison CHILD VS. YOUNG with age group ($\beta = 0.873, p = 0.011$) indicates general age-related differences in approaching the child interlocutor compared to the young interlocutor: The F0 range was larger for young speakers than that of older speakers when addressing the child in comparison to the young interlocutor. These age-related patterns diverge even more when context-related prosodic disambiguation is considered and condition is taken into account. The significant three-way interaction of condition, context comparison CHILD VS. context YOUNG, and age group ($\beta = -0.799, p = 0.018$) indicates that young speakers increased the F0 range in both conditions, brack and nobrack, when addressing the child as compared to the young interlocutor, while older speakers did so only in the brack condition. In the nobrack condition, however, older speakers decreased the F0 range, resulting in an enhanced difference between the conditions when addressing the child as compared to the young interlocutor. Model predictions for F0 range on Name2 by condition, context, and age group are displayed in Figure 8.

3.4.2 Final lengthening on Name2

Results for final lengthening on Name2 are reported from a reduced model.[6] Several effects were statistically significant (see Table 7 and link in Section 2.6). The statistically significant main effect of condition ($\beta = 5.071, p < 0.0001$) confirms that final lengthening was used for the disambiguation of brack and nobrack on Name2 by speakers of both age groups: Final lengthening was increased in the brack condition (the final vowel of Name2 span about 45% of the total duration of Name2) compared to the nobrack condition (the final vowel span about 35% of the total name duration). Regarding age-related differences in prosodic disambiguation and situational (in)dependence, the three-way interaction between

---

[5] The model can be found on https://osf.io/fc8nz.

[6] The model can be found on https://osf.io/fc8nz.

Table 6: Selected model estimates and 95% confidence intervals of the fixed effects for F0 range on Name2 including main effect of condition and main effect and interactions of age group. * $p < 0.05$; ** $p < 0.01$.

Predictor	Estimate	95% CI
Intercept	7.097**	(6.370, 7.824)
condition	3.040**	(2.613, 3.468)
condition:age group	−0.345	(−1.200, 0.510)
child vs. young:age group	0.873*	(0.121, 1.626)
elderly vs. young:age group	0.573	(−0.413, 1.559)
non-native vs. young:age group	0.636	(−0.384, 1.655)
noise vs. young:age group	0.642	(−0.351, 1.635)
condition:child vs. young:age group	−0.779*	(−1.419, −0.139)
condition:elderly vs. young:age group	−0.684	(−1.524, 0.156)
condition:non-native vs. young:age group	−0.306	(−0.968, 0.356)
condition:noise vs. young:age group	−0.442	(−1.193, 0.309)

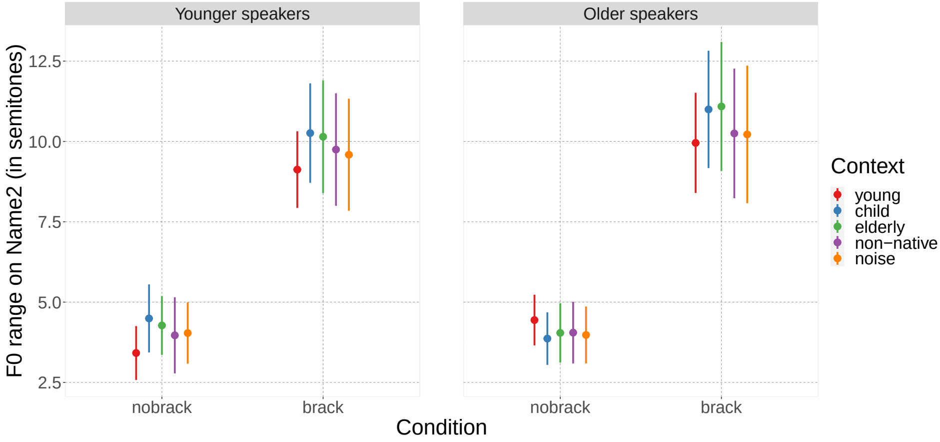


Figure 8: Model predictions for F0 range on Name2 (y-axis) divided by age group (younger speakers left panel, older speakers right panel), condition (x-axis), and context (colour). Whiskers show 95% confidence intervals.

Table 7: Selected model estimates and 95% confidence intervals of the fixed effects for final lengthening on Name2 including main effect of condition and main effect and interactions of age group. * $p < 0.05$; ** $p < 0.01$.

Predictor	Estimate	95% CI
Intercept	40.813*	(39.477, 42.149)
condition	5.071*	(4.284, 5.858)
condition:age group	−0.248	(−1.822, 1.325)
child vs. young:age group	0.766	(−0.584, 2.116)
elderly vs. young:age group	0.943	(−0.399, 2.286)
non-native vs. young:age group	0.880	(−0.467, 2.227)
noise vs. young:age group	1.374	(−0.368, 3.115)
condition:child vs. young:age group	−1.811**	(−3.161, −0.462)
condition:elderly vs. young:age group	−1.939**	(−3.281, −0.596)
condition:non-native vs. young:age group	−1.455	(−3.035, 0.125)
condition:noise vs. young:age group	−0.536	(−1.898, 0.827)

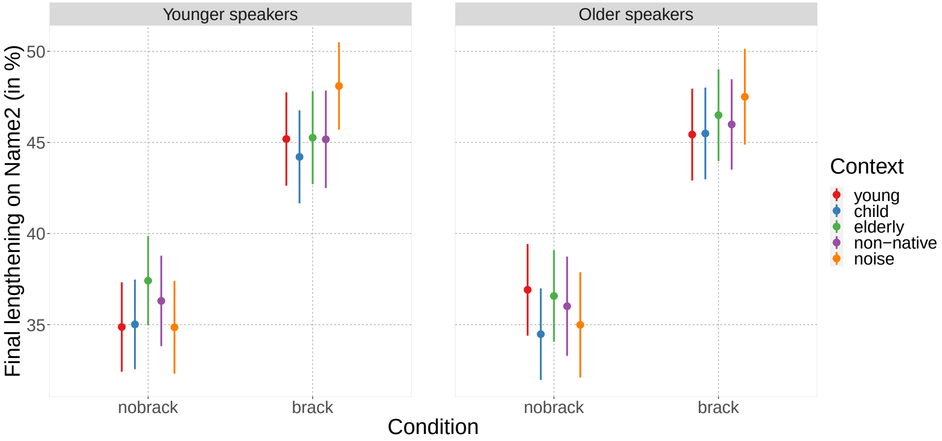


Figure 9: Model predictions for final lengthening on Name2 (y-axis) divided by age group (younger speakers left panel, older speakers right panel), condition (x-axis), and context (colour). Whiskers show 95% confidence intervals.

condition, the context comparison CHILD VS. YOUNG, and age group ($\beta = -1.811$, $p = 0.009$) indicates that young speakers decreased final lengthening in the brack condition when addressing the child as compared to addressing the young interlocutor, thus decreasing the difference between the conditions. On the contrary, older speakers decreased final lengthening for the same context comparison in the nobrack condition, thus increasing the difference between the conditions. An additional three-way interaction between condition, ELDERLY VS. YOUNG and age group ($\beta = -1.939, p = 0.005$) indicates that young speakers increased final lengthening in the nobrack condition when addressing the elderly as compared to the young interlocutor. That is, they reduced the difference between the conditions in context ELDERLY, compared to context YOUNG. Older speakers showed a different behaviour: They increased final lengthening when addressing the elderly as compared to the young interlocutor in the brack condition, thus enhancing the difference between the conditions in context ELDERLY. Model predictions for final lengthening on Name2 by condition, context, and age group are displayed in Figure 9.

3.4.3 Pause after Name2

Since the random effects structure of the model analysing pause after Name2 could not be reduced without a significant drop in model fit, results are reported from the maximal model. None of the effects were statistically significant (see Table 8 and link in Section 2.6).

Table 8: Selected model estimates and 95% confidence intervals of the fixed effects for pause after Name2 including main effects and interactions of age group. * $p < 0.05$; ** $p < 0.01$.

Predictor	Estimate	95% CI
Intercept	19.398**	(14.827, 23.968)
age group	−0.019	(−9.160, 9.122)
CHILD VS. YOUNG:age group	−1.560	(−10.505, 7.386)
ELDERLY VS. YOUNG:age group	−1.202	(−10.113, 7.709)
NON-NATIVE VS. YOUNG:age group	−1.048	(−10.001, 7.906)
NOISE VS. YOUNG:age group	16.472	(−21.986, 54.930)

4 Discussion

In the current study, we compared the use of prosodic cues produced to disambiguate the internal grouping of coordinated three-name sequences (coordinates) in two conditions, that is, without and with internal grouping of the first two names (nobrack and brack, respectively) between two age groups: young (19–34 years) and older (61–80 years) speakers of German. We focused our analysis on the three prosodic cues F0 range, final lengthening, and pause on/after Name1 and Name2. As age affects the stability and variability of tonal and durational features in general, we tested for potential age effects on the modulation of the three prosodic cues for structural disambiguation. Furthermore, we explored whether the situational (in)dependence of disambiguating prosody differs between younger and older speakers, considering their prosodic adaptation to varying contexts. To this end, in both age groups, we elicited coordinates by means of a referential communication task with five contexts: addressing a young adult, a child, an elderly adult, a young non-native adult, and the young adult with background noise.

Looking at the data, we note two things: First of all, descriptively, younger and older speakers produced the three prosodic cues overall quite similarly for prosodic disambiguation and even in the different contexts. This visual observation receives support from the statistical models: For none of the prosodic cues, did the statistical models reveal a main effect of age group. That is, for the use of prosodic cues to mark the internal grouping of coordinates, our data do not provide evidence for a general age-related effect. Second, despite the similarity of the produced prosodic cues, the productions of the older group of speakers are more variable than those of the younger ones, an effect that is evident in larger standard deviations and confidence intervals of the model estimates and the raw data. Increased variability with increased age regarding F0 and durational values is in line with findings of previous studies (Scukanec et al. 1992, 1996, Lortie et al. 2015, Santos et al. 2021, among others).

Regarding our first research question, whether older compared to younger speakers show a more extreme use of F0 range, final lengthening, and pause on Name1 and Name2 to mark the internal grouping, our data do not provide evidence for age-related increases in cue use. In absolute measures, though, older speakers produced longer utterances and longer final vowels on Name1 and Name2 than young speakers, which corresponds to a slower speaking rate since all productions had the same number of syllables. A slower speaking rate is in line with previous findings in the literature (Kemper et al. 1995, Scukanec et al. 1996, Harnsberger et al. 2008, Barnes 2013, Smiljanic & Gilbert 2017, Dimitrova et al. 2018, Tuomainen & Hazan 2018, Hazan et al. 2019, Tuomainen et al. 2019, 2021). Nevertheless, independent of age, speakers in both age groups marked the internal grouping globally in line with the Proximity/Similarity model (Kentner & Féry 2013) using all three cues investigated: In the brack condition, on Name1, speakers of both age groups produced a smaller F0 range and less final lengthening compared to the nobrack condition. This is considered a weakening of the prosodic boundary indicating the sisterhood of the neighbouring element (i.e., Name2 in this case) by Kentner & Féry (2013). On Name2, this pattern was reversed: In the brack condition, speakers of both age groups increased the F0 range and the lengthening of the final segment compared to the nobrack condition and, additionally, inserted a pause after Name2 in the brack condition. This increase of prosodic cues is considered a strengthening of a prosodic boundary (Kentner & Féry 2013). For none of the prosodic cues was the interaction between age group and condition statistically significant. We, thus, did not find support for age-related more extreme use of disambiguating prosodic cues. Across both age groups, the results of the perception checks confirmed that the internal grouping was produced successfully, as the conditions could reliably be recovered by naïve listeners. Only about 4% of the data in each age group led to misunderstandings. That is, despite the variability in the data, speakers of both age groups produced the disambiguating prosodic cues in such a clear way that listeners could correctly resolve the underlying syntactic structure.

Regarding our second research question, whether young and older speakers differ in adapting their use of prosodic cues when addressing varying interlocutors, our data show substantial similarities across age groups. For several model predictions, the estimated means of the non-baseline contexts within one condition deviate in the same direction from the young baseline context in both age groups (cf. brack in Figures 7 and 8). This also explains why only few interactions of context, condition, and age group revealed statistical significance. Nevertheless, there are slight differences between the age groups regarding their adaptations. We will focus our discussion on statistically significant three-way-interactions of age groups, contexts, and condition, as we are mainly interested in the interplay of all three factors.

The two age groups diverged most strongly when addressing a child as compared to a young interlocutor: On Name2, the older speakers produced larger F0 ranges for the child compared to the young interlocutor in condition brack and smaller F0 ranges along with decreased final lengthening in condition nobrack, thus increasing the difference between conditions when addressing the child. Younger speakers, however, rather slightly decreased the difference between brack and nobrack when addressing the child as they reduced final lengthening in the brack condition. This enhanced difference between conditions in older speakers can be interpreted as more adaptation to the child interlocutor in older than in younger speakers. Such an enhanced difference between conditions in older compared to younger speakers also holds true for the context with the elderly interlocutor. Here, the older speakers slightly increased the difference between the conditions by means of an increase in final lengthening on Name2 in the brack condition while the young speakers showed the reverse pattern: They decreased the difference in final lengthening between the conditions by increasing final lengthening in nobrack. Interestingly, from the viewpoint of disambiguation, speakers in both age groups produced a stronger distinction between conditions when addressing their peer compared to addressing a non-age-matched interlocutor: young speakers addressing the young interlocutor and older speakers addressing the elderly interlocutor. We are not aware of any similar findings in the literature. Yet, despite being statistically significant, these differences in adaptation between age groups were in fact quite small in absolute terms, and did not affect the disambiguation of coordinates, as revealed by the perception check. Together with the large variability in the productions of the older speakers (cf. larger 95% confidence intervals in the Figures with model predictions than for younger speakers), it is questionable whether the effects in the child and elderly contexts compared to the young context are reproducible in the same manner in future studies.

In the remaining two contexts (non-native interlocutor and speech in noise), our data did not demonstrate evidence for differences between the age groups. Given this and given the fact that any context differences across groups did not impact on disambiguation of coordinates in general, regarding our second research question, our data speak in favour of situational independence in both age groups (Schafer et al. 2000, Kraljic & Brennan 2005, Speer et al. 2011). Models of situational independence assume that disambiguating prosody is realised automatically as part of the production process on the side of the speaker and is therefore largely independent of the presence or absence of a listener, the type of listener, or the situational setting. As such, it seems plausible that disambiguating prosody is also independent of the age of the speaker. Our data add to the literature on the effects of different types of interlocutors and the absence/presence of noise on the use of disambiguating prosodic cues the dimension of speaker age. The findings show that situational independence in production of disambiguating prosody holds for older speakers, too, and that prosody production is a stable automatic part of the production process also in older speakers.

Thus, whereas age has frequently been shown to affect other areas of language production (i.e., word-finding abilities, increased phonetic variability, or altered acoustic characteristics), it does not seem to have a (listener-relevant) impact on production of prosodic cues in ambiguous structures. This is in line with an observation by Lortie et al. (2015) regarding a more variable voice in older speakers that did not interact with the ability to control fundamental frequency (participants in their study were asked to produce normal, low, and high frequency voice in sustained vowels). In this sense, our study provides evidence that one important part of the prosody-syntax interface is not modulated by age effects: the use of the prosodic cues F0 range, final lengthening, and pause for disambiguation of structurally ambiguous coordinates. Our findings on prosody production in older adults are also of importance in the larger context of investigating linguistic prosody in populations with acquired language and communication disorders resulting from brain lesions (i.e., aphasia or right-hemisphere brain lesions), since participants in these studies are usually older than the typical age groups covered in most studies on healthy prosody processing.

In summary, our data confirm the well-known general age-related changes in absolute durational measures. However, when it comes to the use of tonal and durational prosodic cues to disambiguate the underlying syntactic structure, older speakers modulated duration and F0 range similarly to younger speakers with, if at all, only minimal differences between the the age groups of speakers in our sample. The finding of limited adaptation to different interlocutors favours models of situational independence of disambiguating prosody across both age groups and shows that production of disambiguating prosody at the prosody-syntax interface is unaffected by age.

5 Conclusion

In conclusion, young and older speakers in our production study globally marked the internal grouping of coordinated name sequences using F0 range, final lengthening, and pause in a similar way. The modulation of disambiguating prosodic cues seems to be independent of age-related changes in absolute durations. Across both age groups, the use of prosodic cues to resolve the ambiguity in the internal structure of coordinates dominated in comparison to possible prosodic accommodations to the contexts, which we interpret as evidence for situational independence of disambiguating prosody. Prosodic disambiguation thus turns out to be a stable automatic part of the production process, regardless of speaker age.

Acknowledgements

We would like to thank L. Junack for her help with acquisition and pre-processing of the data, as well as for setting up the perception check for the data of the older speakers.

Funding information

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), SFB 1287 project B01 (project no. 317633480).

References

Barnes, Daniel R. 2013. *Age-related changes to the production of linguistic prosody*. Purdue University. (Doctoral dissertation). https://docs.lib.purdue.edu/open_access_theses/17.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious mixed models. *arXiv: Methodology*. DOI: 10.48550/arXiv.1506.04967.

Biersack, Sonja, Vera Kempe & Lorna Knapton. 2005. Fine-tuning speech registers: A comparison of the prosodic features of child-directed and foreigner-directed speech. In *Proceedings of InterSpeech*, 2401–2404. DOI: 10.21437/Interspeech.2005-46.

Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer*. http://www.praat.org/.

Burke, Deborah M. & Meredith A. Shafto. 2004. Aging and language production. *Current Directions in Psychological Science* 13(1). 21–24. DOI: 10.1111/j.0963-7214.2004.01301006.x.

DePaulo, Bella M. & Lerita M. Coleman. 1986. Talking to children, foreigners, and retarded adults. *Journal of Personality and Social Psychology* 51(5). 945–959. DOI: 10.1037/0022-3514.51.5.945.

Dimitrova, Snezhina, Bistra Andreeva, Christoph Gabriel & Jonas Grünke. 2018. Speaker age effects on prosodic patterns in Bulgarian. In *Proceedings of Speech Prosody*, 709–713. DOI: 10.21437/SpeechProsody.2018-144.

Dromey, Christopher & Sarah Scott. 2016. The effects of noise on speech movements in young, middle-aged, and older adults. *Speech, Language and Hearing* 19(3). 131–139. DOI: 10.1080/2050571X.2015.1133757.

Gollrad, Anja, Esther Sommerfeld & Frank Kügler. 2010. Prosodic cue weighting in disambiguation: Case ambiguity in German. In *Proceedings of Speech Prosody*, 165. https://www.isca-speech.org/archive/pdfs/speechprosody_2010/gollrad10_speechprosody.pdf.

Goy, Huiwen, David N. Fernandes, M. Kathleen Pichora-Fuller & Pascal van Lieshout. 2013. Normative voice data for younger and older adults. *Journal of Voice* 27(5). 545–555. DOI: 10.1016/j.jvoice.2013.03.002.

Hansen, Marie, Clara Huttenlauch, Carola de Beer, Isabell Wartenburger & Sandra Hanne. 2022. Individual differences in early disambiguation of prosodic grouping. *Language and Speech* OnlineFirst. DOI: 10.1177/00238309221127374.

Harnsberger, James D., Rahul Shrivastav, William S. Brown Jr, Howard Rothman & Harry Hollien. 2008. Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice* 22(1). 58–69. DOI: 10.1016/j.jvoice.2006.07.004.

Hazan, Valerie, Outi Tuomainen, Jeesun Kim & Chris Davis. 2019. The effect of visual cues on speech characteristics of older and younger adults in an interactive task. In *Proceedings of the 19th ICPhS*, 815–819. https://www.assta.org/proceedings/ICPhS2019/papers/ICPhS_864.pdf.

Heister, Julian, Kay-Michael Würzner, Johannes Bubenzer, Edmund Pohl, Thomas Hanneforth, Alexander Geyken & Reinhold Kliegl. 2011. dlexDB: Eine lexikalische Datenbank für die psychologische und linguistische Forschung. In *Psychologische Rundschau*, vol. 62, 10–20. Göttingen: Hogrefe. DOI: 10.1026/0033-3042/a000029.

Holzgrefe-Lang, Julia, Caroline Wellmann, Caterina Petrone, Romy Räling, Hubert Truckenbrodt, Barbara Höhle & Isabell Wartenburger. 2016. How pitch change and final lengthening cue boundary perception in German: Converging evidence from ERPs and prosodic judgements. *Language, Cognition and Neuroscience* 31(7). 904–920. DOI: 10.1080/23273798.2016.1157195.

Howell, David C., Marylène Rogier, Vincent Yzerbyt & Yves Bestgen. 1998. *Statistical methods in human sciences*. New York: Wadsworth.

Huttenlauch, Clara, Carola de Beer, Sandra Hanne & Isabell Wartenburger. 2021. Production of prosodic cues in coordinate name sequences addressing varying interlocutors. *Laboratory Phonology* 12(1). 1–31. DOI: 10.5334/labphon.221.

Jessen, Michael, Olaf Köster & Stefan Gfroerer. 2003. Effect of increased vocal effort on average and range of fundamental frequency in a sample of 100 German-speaking male subjects. In *Proceedings of the 15th ICPhS*, 1623–1626. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_1623.html.

Junqua, Jean-Claude. 1996. The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication* 20(1). 13–22. DOI: 10.1016/S0167-6393(96)00041-6.

Kemper, Susan, Patrice Ferrell, Tamara Harden, Andrea Finter-Urczyk & Catherine Billington. 1998. Use of elderspeak by young and older adults to impaired and unimpaired listeners. *Aging, Neuropsychology, and Cognition* 5(1). 43–55. DOI: 10.1076/anec.5.1.43.22.

Kemper, Susan, Dixie Vandeputte, Karla Rice, Him Cheung & Julia Gubarchuk. 1995. Speech adjustments to aging during a referential communication task. *Journal of Language and Social Psychology* 14(1). 40–59. DOI: 10.1177/0261927X95141003.

Kentner, Gerrit & Caroline Féry. 2013. A new approach to prosodic grouping. *The Linguistic Review* 30. 277–311. DOI: 10.1515/tlr-2013-0009.

Knoll, Monja A., Melissa Johnstone & Charlene Blakely. 2015. Can you hear me? Acoustic modifications in speech directed to foreigners and hearing-impaired people. In *Proceedings of InterSpeech*, 2987–2990. DOI: 10.21437/Interspeech.2015-618.

Knoll, Monja A. & Lisa Scharrer. 2007. Acoustic and affective comparisons of natural and imaginary infant-, foreigner- and adult-directed speech. In *Proceedings of InterSpeech*, 1414–1417. DOI: 10.21437/Interspeech.2007-29.

Knoll, Monja A., Lisa Scharrer & Alan Costall. 2011. "Look at the shark": Evaluation of student- and actress-produced standardised sentences of infant- and foreigner-directed speech. *Speech Communication* 53(1). 12–22. DOI: 10.1016/j.specom.2010.08.004.

Kraljic, Tanya & Susan E. Brennan. 2005. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology* 50(2). 194–231. DOI: 10.1016/j.cogpsych.2004.08.002.

Lange, Kristian, Simone Kühn & Elisa Filevich. 2015. "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PloS ONE* 10(6). e0130834. DOI: 10.1371/journal.pone.0130834.

Lin, Yuri, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman & Slav Petrov. 2012. Syntactic annotations for the Google Books Ngram Corpus. In *Proceedings of the 50th annual meeting of the Association for Computational Linguistics*, 169–174. https://aclanthology.org/P12-3029.pdf.

Lombard, Étienne. 1911. Le signe de l'élévation de la voix. *Annales des Maladies de l'Oreille et du Larynx* 37. 101–109.

Lortie, Catherine L., Mélanie Thibeault, Matthieu J. Guitton & Pascale Tremblay. 2015. Effects of age on the amplitude, frequency and perceived quality of voice. *AGE* 37(6). 1–24. DOI: 10.1007/s11357-015-9854-1.

Markó, Alexandra & Judit Bóna. 2010. Fundamental frequency patterns: The factors of age and speech type. In *Proceedings of the "Workshop sociophonetics, at the crossroads of speech variation, processing and communication"*, 45–48. https://edizioni.sns.it/wp-content/uploads/2013/12/Sociophonetics_12_MARKO.pdf.

Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2). 314–324. DOI: 10.3758/s13428-011-0168-7.

Nasreddine, Ziad S., Natalie A. Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L. Cummings & Howard Chertkow. 2005. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society* 53(4). 695–699. DOI: 10.1111/j.1532-5415.2005.53221.x.

Neurobehavioural Systems. 2018. Version 20.1. https://www.neurobs.com/.

Olusanya, Bolajoko O., Adrian C. Davis & Howard J. Hoffman. 2019. Hearing loss grades and the international classification of functioning, disability and health. *Bulletin of the World Health Organization* 97(10). 725–728. DOI: 10.2471/BLT.19.230367.

Peters, Benno, Klaus J. Kohler & Thomas Wesener. 2005. Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In Klaus J. Kohler, Felicitas Kleber & Benno Peters (eds.), *Prosodic structures in German spontaneous speech*, 143–184. Kiel: Arbeitsberichte (AIPUK). https://www.ipds.uni-kiel.de/kjk/pub_exx/aipuk35a/aipuk35a_6.pdf.

Petrone, Caterina, Hubert Truckenbrodt, Caroline Wellmann, Julia Holzgrefe-Lang, Isabell Wartenburger & Barbara Höhle. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61. 71–92. DOI: 10.1016/j.wocn.2017.01.002.

Piazza, Giorgio, Clara D. Martin & Marina Kalashnikova. 2021. The acoustic features and didactic function of foreigner directed speech: A literature review. DOI: 10.31234/osf.io/scnke.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Rodriguez-Cuadrado, Sara, Cristina Baus & Albert Costa. 2018. Foreigner talk through word reduction in native/non-native spoken interactions. *Bilingualism: Language and Cognition* 21(2). 419–426. DOI: 10.1017/S1366728917000402.

Rojas, Sandra, Elaina Kefalianos & Adam Vogel. 2020. How does our voice change as we age? A systematic review and meta-analysis of acoustic and perceptual voice data from healthy adults over 50 years of age. *Journal of Speech, Language, and Hearing Research* 63(2). 533–551. DOI: 10.1044/2019_jslhr-19-00099.

Santos, Aline Oliveira, Juliana Godoy, Kelly Silverio & Alcione Brasolotto. 2021. Vocal changes of men and women from different age decades: An analysis from 30 years of age. *Journal of Voice* OnlineFirst. DOI: 10.1016/j.jvoice.2021.06.003.

Schafer, Amy J., Shari R. Speer, Paul Warren & S. David White. 2000. Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research* 29(2). 169–182. DOI: 10.1023/A:1005192911512.

Schubö, Fabian & Sabine Zerbian. 2023. The patterns of pre-boundary lengthening in German. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 1–34. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777526.

Scukanec, Gail P., Linda Petrosino & Roger D. Colcord. 1996. Age-related differences in acoustical aspects of contrastive stress in women. *Folia Phoniatrica et Logopaedica* 48(5). 231–239. DOI: 10.1159/000266414.

Scukanec, Gail P., Linda Petrosino & Michael P. Rastatter. 1992. Fundamental frequency variability in elderly women during production of stressed and unstressed words. *Perceptual and Motor Skills* 74(3_suppl). 1091–1095. DOI: 10.2466/pms.1992.74.3c.1091.

Smiljanic, Rajka & Rachael C. Gilbert. 2017. Acoustics of clear and noise-adapted speech in children, young, and older adults. *Journal of Speech, Language, and Hearing Research* 60(11). 3081–3096. DOI: 10.1044/2017_JSLHR-S-16-0130.

Smith, Caroline L. 2007. Prosodic accommodation by French speakers to a nonnative interlocutor. In *Proceedings of the 16th ICPhS*, 1081–1084. http://www.icphs2007.de/conference/Papers/1116/1116.pdf.

Speer, Shari R., Paul Warren & Amy J. Schafer. 2011. Situationally independent prosodic phrasing. *Laboratory Phonology* 2(1). 35–98. DOI: 10.1515/labphon.2011.002.

Summers, W. Van, David B. Pisoni, Robert H. Bernacki, Robert I. Pedlow & Michael A. Stokes. 1988. Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America* 84(3). 917–928. DOI: 10.1121/1.396660.

Tauber, Sarah K., Lori E. James & Paula M. Noble. 2010. The effects of age on using prosody to convey meaning and on judging communicative effectiveness. *Psychology and Aging* 25(3). 702–707. DOI: 10.1037/a0019266.

Thimm, Caja, Ute Rademacher & Lenelis Kruse. 1998. Age stereotypes and patronizing messages: Features of age-adapted speech in technical instructions to the elderly. *Journal of Applied Communication Research* 26(1). 66–82. DOI: 10.1080/00909889809365492.

Tucker, Benjamin V., Catherine Ford & Stephanie Hedges. 2021. Speech aging: Production and perception. *Wiley Interdisciplinary Reviews: Cognitive Science* 12. e1557. DOI: 10.1002/wcs.1557.

Tuomainen, Outi & Valerie Hazan. 2018. Investigating clear speech adaptations in spontaneous speech produced in communicative settings. In Mária Gósy & Tekla Etelka Gráczi (eds.), *Challenges in analysis and processing of spontaneous speech*, 9–25. Budapest: Hungarian Academy of Sciences. DOI: 10.18135/CAPSS.9.

Tuomainen, Outi, Valerie Hazan & Linda Taschenberger. 2019. Speech communication in background noise: Effects of aging. In *Proceedings of the 19th ICPhS*, 805–809. https://www.assta.org/proceedings/ICPhS2019/papers/ICPhS_854.pdf.

Tuomainen, Outi, Linda Taschenberger, Stuart Rosen & Valerie Hazan. 2021. Speech modifications in interactive speech: Effects of age, sex and noise type. *Philosophical Transactions of the Royal Society B: Biological Sciences* 377(1841). 20200398. DOI: 10.1098/rstb.2020.0398.

Turk, Alice E., Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. In Stefan Sudhoff, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.), *Methods in empirical prosody research*, 1–28. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110914641.1.

Uther, Maria, Monja A. Knoll & Denis Burnham. 2007. Do you speak E-NG-LI-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication* 49(1). 2–7. DOI: 10.1016/j.specom.2006.10.003.

Wellmann, Caroline, Julia Holzgrefe-Lang, Hubert Truckenbrodt, Isabell Wartenburger & Barbara Höhle. 2023. Developmental changes in prosodic boundary cue perception in German-learning infants. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.), *Prosodic boundary phenomena*, 119–156. Berlin: Language Science Press. DOI: 10.5281/zenodo.7777532.

Zollinger, Sue Anne & Henrik Brumm. 2011. The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* 148(11). 1173–1198. DOI: 10.1163/000579511X605759.

Name index

Aasland, Wendi A., 120, 145
Abramson, Arthur S., 49
Ahn, Suzy, 43, 51
Alves, Mariane Antero, 43
Ames, Elinor W., 135
Asu, Eva Liina, 95
Audacity Team, 46
Avelino, Heriberto, 51

Bar, Moshe, 110
Barnes, Daniel R., 161, 162, 183
Barnes, Jonathan A., 67
Bates, Douglas, 16, 97, 172
Baum, Shari R., 120, 145
Baumann, Stefan, 109, 110
Beach, Cheryl M., 147
Beckman, Mary E., 37–40, 44
Bell, Alan, 52, 54
Beňuš, Štefan, 52
Berkovits, Rochele, 6, 23, 37, 63, 89, 107
Best, Catherine C., 120
Bhatara, Anjali, 134
Biersack, Sonja, 164
Bijeljac-Babic, Ranka, 120, 134, 135
Biron, Tirza, 107
Blumstein, Sheila E., 51
Bodur, Kübra, 38
Boersma, Paul, 16, 53, 95, 132, 158, 170
Bois, John W. Du, 89
Bombien, Lasse, 62
Bóna, Judit, 161
Bosch, Laura, 141
Breen, Mara, 89
Brennan, Susan E., 163, 185
Browman, Catherine P., 4
Brumm, Henrik, 164, 165
Buhmann, Jeska, 89
Burke, Deborah M., 160
Burnham, Denis, 120, 122
Butcher, Andy, 132, 143
Byrd, Dani, 2, 4–6, 23, 37, 62, 63

Câmara JR, Joaquim Mattoso, 41
Cambier-Langeveld, Tina, 3, 5, 6, 25, 68
Cantoni, Maria Mendes, 40, 48
Carlson, Katy, 65
Chafe, Wallace, 88
Chater, Nick, 87
Cho, Taehong, 4, 5, 23, 35–38, 41–43, 47, 51, 61–63, 65, 69, 70
Christiansen, Morten H., 87
Christodoulides, George, 88, 91, 110
Clark, Andy, 110
Clopper, Cynthia G., 69
Cole, Jennifer, 36, 38, 88, 91–94, 110
Coleman, Lerita M., 164
Collier, René, 120, 143
Cooper, William E., 89, 96
Couper-Kuhlen, Elizabeth, 89
Cristià, Alejandrina, 120, 121, 141, 145
Cristófaro-Silva, Thaïs, 40, 41, 46
Crosswhite, Katherine M., 41

Crystal, Thomas H., 46, 51
Cutler, Anne, 91

D’Andrade, Ernesto, 40, 41
Dahan, Delphine, 87, 88
Davidson, Lisa, 44
Davies, Mark, 48, 52
de Jong, Kenneth J., 46
de Jong, Nivja H., 51
de la Cruz-Pavía, Irene, 90
de la Mota, Carme, 53
Delattre, Pierre, 7
Delforge, Ann Marie, 44
DePaulo, Bella M., 164
Di Cristo, Albert, 120
Diehl, Randy L., 70
Dilley, Laura, 37
Dimitrova, Snezhina, 161, 183
Dogil, Grzegorz, 7
Dromey, Christopher, 165
Duez, Danielle, 88, 90

Edwards, Jan, 39
Eek, Arvo, 108
Engel, Andreas K., 110
Erelt, Mati, 96
Erickson, Donna, 42
Ernestus, Mirjam, 42

Fernald, Anne, 142, 146
Ferreira, Fernanda, 87, 88
Ferreira, Michael, 48, 52
Féry, Caroline, vi, vii, 10, 12, 91, 107, 122, 157–159, 184
Fisher, Cynthia, 142
Fletcher, Janet, 36, 37
Fon, Janice, 89, 107
Fónagy, Ivan, 47, 51
Fougeron, Cecile, 35, 37, 38, 41, 61
Fowler, Carol A., 52
Frazier, Lyn, v
Friederici, Angela D., 148
Frota, Sonia, 37, 39, 40, 53, 63, 69
Fujisaki, Hiroya, 89

Garellek, Marc, 52, 53, 65
Georgeton, Laurianne, 37
Gilbert, Rachael C., 161, 165, 183
Goldman, Jean-Philippe, 53
Goldman-Eisler, Frieda, 143
Goldstein, Louis, 4
Gollrad, Anja, 158
Gordon, Matthew, 41
Götz, Antonia, 120
Gout, Ariel, 120
Goy, Huiwen, 161
Green, Clarence, 103
Grice, Martine, 7, 8, 13
Gries, Stefan Thomas, 54
Gussenhoven, Carlos, 37

Hall, Tracy Alan, 7
Halliday, Michael A. K., 88
Hansen, Marie, 159
Harnsberger, James D., 161, 183
Hazan, Valerie, 161, 184
Heister, Julian, 167
Herold, Birgit, 141
Himmelmann, Nikolaus P., 88, 89, 92, 94, 107, 108
Hirose, Keikichi, 89
Hirschberg, Julia Bell, 53, 89
Hirsh-Pasek, Kathy, 120
Hirst, Daniel, 120
Hofhuis, Elise, 6
Höhle, Barbara, 120, 134, 141, 148
Holm, Sture, 139

Holzgrefe-Lang, Julia, 122, 123, 141, 147, 148, 166
Hothorn, Torsten, 16
House, Arthur S., 46, 51
Housum, Jonathan, 52
Howell, David C., 170
Hsu, Chai-Shune, 37
Hualde, José Ignacio, 39, 40, 46, 48
Hunter, Michael A., 135
Huttenlauch, Clara, v–vii, 10, 89, 90, 107, 122, 157–160, 163–167, 169–171

Ishihara, Shinichiro, 91

Jessen, Michael, 165
Johnson, Elizabeth K., 120–122, 143, 145, 146
Jones, Daniel, 21
Jun, Sun-Ah, 36, 37, 39, 44, 65
Junqua, Jean-Claude, 165
Jusczyk, Peter W., 120

Katsika, Argyro, vi, 4, 24, 36, 37, 63, 64, 69
Keating, Patricia A., 36–39, 42, 48, 61, 62, 88
Kemler Nelson, Deborah G., 120
Kemper, Susan, 161, 164, 183
Kentner, Gerrit, vi, vii, 10, 107, 122, 157–159, 184
Kessinger, Rachel H., 51
Keyser, Samuel Jay, 42
Kim, Sahyang, 44, 63
Kingston, John, 70
Klatt, Dennis H., 6
Knoll, Monja A., 164
Koch, Gary G., 102
Kohler, Klaus J., 2–4, 6, 8, 22, 128
Kraljic, Tanya, 163, 185
Krivokapić, Jelena, 36, 52
Krull, Diana, 3, 63
Kügler, Frank, 12
Kuhl, Patricia K., 120
Kuzla, Claudia, 38

Ladd, D. Robert, 39, 40, 88, 89
Ladefoged, Peter, 51, 70
Landis, J. Richard, 102
Lange, Kristian, 170
Langus, Alan, 88, 90
Lavoie, Lisa M., 37, 38, 41, 42, 57
Lehiste, Ilse, 107, 108, 120
Levshina, Natalia, 75
Liberman, Mark Y., 40, 89, 96
Lin, Yuri, 167
Lippus, Pärtel, 93, 104
Lisker, Leigh, 49
Lombard, Étienne, 164
Lortie, Catherine L., 161, 183, 186
Ludusan, Bogdan, 142

Magdics, Klara, 47, 51
Mahrt, Tim, 92–94
Männel, Claudia, 148
Markó, Alexandra, 161
Mateus, Maria Helena, 40, 41
Mathôt, Sebastiaan, 170
Mattock, Karen, 120, 122
Mattys, Sven L., 88
Maurer, Daphne, 120
McAuliffe, Michael, 90
McQueen, James M., 36, 37, 43, 47, 61
McRoberts, Gerald W., 120
Metslang, Helle, 96
Miller, George A., 103
Mo, Yoonsook, 44
Moraes, Joao, 39, 53

Nadeu, Marianna, 41, 46
Nakai, Satsuki, 2, 3, 6, 89, 107
Napoleão de Souza, Ricardo, v, vi, 23
Nash, John C., 98
Nasreddine, Ziad S., 166
Nazzi, Thierry, 120–122, 143
Nespor, Marina, v, 36, 37, 120
Nolan, Francis, 95

Oh, Sejin, 42
Oller, D. Kimbrough, 37, 63
Olusanya, Bolajoko O., 166
Onaka, Akiko, 38
Ordin, Mikhail, 88, 90
Ortega-Llebaria, Marta, 40
Ots, Nele, v, vii, 93, 106

Parrell, Benjamin, 37, 41
Peters, Benno, 4, 13, 108, 136, 142, 147, 158
Peters, Jörg, 7
Petrone, Caterina, 1, 4, 8, 10, 12–14, 63, 89–91, 107, 120, 122, 128, 158, 171
Piazza, Giorgio, 164
Pierrehumbert, Janet B., 37–39, 53, 88, 89, 96
Plag, Ingo, 40
Polka, Linda, 120, 122
Price, Patti J., 120
Prieto, Pilar, 37, 39, 40, 48, 69
Prince, Alan, 40

R Core Team, 16, 54, 97, 171
Ramanarayanan, Vikram, 52
Rao, Rajiv, 63
Redi, Laura, 53
Riesberg, Sonja, 88, 91, 92, 94, 107, 108, 110
Rietveld, Antonius C. M., 37
Riggs, Daylen, 5, 63
Rodriguez-Cuadrado, Sara, 164
Roettger, Timo, 41
Rojas, Sandra, 161
Roseano, Paolo, 69
Rosenfelder, Ingrid, 53

Saltzman, Elliot, 2, 4, 23
Sanderman, Angelien A., 120
Santos, Aline Oliveira, 161, 183
Sawusch, James R., 48
Schafer, Amy J., 157, 163, 185
Scharrer, Lisa, 164
Scheer, Tobias, 67
Schild, Ulrike, 88
Schipke, Christine S., 148
Schmitz, Michaela, 120, 124, 136
Schubö, Fabian, v, vi, 2, 4, 12, 25, 63, 89, 107, 122, 158
Schuetze-Coburn, Stephan, 89
Scott, Donia R., 120
Scott, Sarah, 165
Scukanec, Gail P., 161, 162, 183
Sebastián-Gallés, Núria, 141
Ségéral, Philippe, 67
Seidl, Amanda, 120–122, 133, 136, 141, 143, 145, 146
Selkirk, Elisabeth, 7, 36, 120
Seo, Jungyun, 5, 6, 23, 24
Shafto, Meredith A., 160
Shattuck-Hufnagel, Stefanie, 3–6, 10, 23–25, 36, 37, 39, 40, 53, 63, 64, 69, 88
Shukla, Mohinish, 120
Silbert, Lauren J., 88
Silverman, Kim E. A., 2, 4, 6, 8, 22, 23
Šimko, Juraj, 52
Simon, Anne Catherine, 91

Simon, Thomas, 142, 146
Singmann, Henrik, 54
Skoruppa, Katrin, 120, 134
Smiljanic, Rajka, 69, 161, 165, 183
Smith, Caroline L., 164
Snedeker, Jesse, 12
Soderstrom, Melanie, 143
Sorensen, John M., 89, 96
Speer, Shari R., vii, 163, 185
Steinhauer, Karsten, 147
Stevens, Kenneth N., 42
Strangert, Eva, 143
Streeter, Lynn A., 120
Summers, W. Van, 165

Tabain, Marija, 37
Talkin, David, 37, 38
Taremaa, Piia, v, vii, 93, 106
Tauber, Sarah K., 162
Tees, Richard C., 120, 122
Terken, Jacques, 143
Thimm, Caja, 164
Thorsen, Nina Grønnum, 89
Tokura, Hisayo, 142
Torreira, Francisco, 42
Trouvain, Jürgen, 90
Truckenbrodt, Hubert, 7, 8, 120, 122, 128
Trueswell, John, 12
Tucker, Benjamin V., 161
Tuomainen, Outi, 161, 165, 184
Turk, Alice E., 3–6, 10, 16, 23–25, 36, 37, 39, 40, 48, 63, 64, 69, 170

Ulbrich, Christiane, 108
Uther, Maria, 164

Vaissière, Jacqueline, 2, 120
van de Weijer, Joost, 146
van Ommen, Sandrien, 122, 131, 143, 145, 146
Varadhan, Ravi, 98
Vigário, Marina, 40
Vogel, Irene, v, 36, 37, 39, 120

Wagner, Michael, 10, 90
Weenink, David, 16, 53, 95, 132, 158, 170
Wellmann, Caroline, v–vii, 10, 89, 90, 107, 110, 119, 122, 123, 125, 131, 132, 141, 142, 144, 166
Wempe, Ton, 51
Werker, Janet F., 120, 122
Whalen, Douglas H., 49
White, Laurence, 2, 3, 23, 24, 38, 63, 88
Wiese, Richard, 7
Wightman, Colin W., 37, 89, 107
Williams, Briony, 7
Winitz, Harris, 51
Winter, Bodo, 109, 110

Xu, Yi, 48

Yeung, H. Henny, 120
Yuan, Jiahong, 89, 96

Zerbian, Sabine, v, vi, 2, 63, 89, 107, 122, 158
Zhang, Xinting, 120
Zollinger, Sue Anne, 164, 165

Prosodic boundary phenomena

In spoken language comprehension, the hearer is faced with a more or less continuous stream of auditory information. Prosodic cues, such as pitch movement, pre-boundary lengthening, and pauses, incrementally help to organize the incoming stream of information into prosodic phrases, which often coincide with syntactic units. Prosody is hence central to spoken language comprehension and some models assume that the speaker produces prosody in a consistent and hierarchical fashion. While there is manifold empirical evidence that prosodic boundary cues are reliably and robustly produced and effectively guide spoken sentence comprehension across different populations and languages, the underlying mechanisms and the nature of the prosody-syntax interface still have not been identified sufficiently. This is also reflected in the fact that most models on sentence processing completely lack prosodic information.

This edited book volume is grounded in a workshop that was held in 2021 at the annual conference of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS). The five chapters cover selected topics on the production and comprehension of prosodic cues in various populations and languages, all focusing in particular on processing of prosody at structurally relevant prosodic boundaries. Specifically, the book comprises cross-linguistic evidence as well as evidence from non-native listeners, infants, adults, and elderly speakers, highlighting the important role of prosody in both language production and comprehension.