Individual Differences in the Production and Perception of Prosodic Boundaries in American English

by

Jiseung Kim

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in The University of Michigan
2020

Doctoral Committee:
  Professor Patrice Speeter Beddor, Co-chair
  Associate Professor Jelena Krivokapić, Co-chair
  Professor Julie Boland
  Professor Jennifer S. Cole, Northwestern University

Jiseung Kim

jiseungk@umich.edu

ORCID iD: 0000-0002-5612-8201

ACKNOWLEDGEMENTS

My sincere and heartfelt gratitude goes to my advisors Pam Beddor and Jelena Krivokapić. It is an understatement to say that this dissertation would not have been possible without your unending support and generosity. If I am ever proud of my work, it is because of your guidance and education. You have never failed to surprise me with your dedication, patience, and wisdom. I feel truly privileged to have been one of your advisees. Your work ethic and commitment to academic rigor have constantly inspired me. I could not have asked for a better pair of advisors than you, and I am forever honored and grateful to call you my mentors.

Thanks to my wonderful committee members, Julie Boland and Jennifer Cole. I am deeply grateful for your extremely valuable input and thoughtful advice. I am lucky to have such an amazing group of scholars as my committee that only helped improve this research. Any shortcomings, mistakes, and limitations found in this dissertation are solely my own.

There are many others I am indebted to and thankful for. I received financial supports from the Linguistics Department, Rackham Graduate School, the Barbour Scholarship, and the International Phonetic Association over the course of this research. Will Styler and Stephen Tobin provided incalculable help in various stages of this research that I am forever grateful for. I would like to thank the participants in this study and my amazing research assistants – Seulgi Shin, Yuna Jeong, and Ella Deaton – for their help. The members of the PhonDi group meetings shared their knowledge and provided feedback whenever I presented my research there and I am thankful for their input.

At the University of Michigan, I have been very fortunate to have such wonderful faculty members – especially Andries Coetzee, Marlyse Baptista, Sally Thomason, Acrisio Pires, Sam Epstein, Ezra Keshet, Robin Queen, Nick Ellis, Steve Abney, Jon Brennan, Natasha Abner, San Duanmu, Savi Namboodiripad, Carmel O'Shannessy, Nick Henriksen, Lorenzo García-Amaya – I feel lucky to have been your student, GSI, RA, and/or catsitter. Also, big thanks to the department's exceptional staff – especially Jennifer Nguyen, Sandie Petee, and Talisha Reviere-Winston – who made my life so much easier.

I would also like to thank Harim Kwon, Ariana Bancu, Moira Saltzman, Marjorie

Herbert, Hayley Heaton, Kate Sherwood for making me feel at home in Ann Arbor. Your friendship means a lot to me. I am also grateful to have known and been friends with many wonderful graduate students during my time at Michigan: Aliaksei Akimenka, Danuta Allen, Fahad Alrashed, Marcus Berger, Felicia Bisnath, Sagan Blue, Rawan Bonais, Dominique Bouavichith, Danielle Burgess, Ian Calloway, Dominique Canning, Tridha Chatterjee, Lauretta Cheng, Yu-Chuan (Lucy) Chiang, Jeonghwa Cho, Justin Craft, Yan Dong, Wil Gonzales, Sujeewa Hettiarachchi, Tamarae Hildebrandt, Alan Ke, Alex Kramer, Chia-wen Lo, Andrew McInnerney, Carrie-Ann Morgan, Will Nediger, L.R. "Nik" Nikolai, Dave Ogden, Mike Opper, Joy Peltier, Cameron Rule, Emily Sabo, Yourdanis Sedarous, Jae-Young Shim, Batia Snir, Theo Stern, Alicia Stevers, Yushi Sugimoto, Tzu-Yun Tung, Stephen Tyndall, Rachel Weissler, Kelly Wright, Jon Yip, Jian Zhu.

Last but not least, 긴 시간 동안 한국에서 큰 힘이 되어준 우리 가족들 특히 엄마랑 아빠 감사합니다.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Theoretical interest in the relation between speech production and perception has led to research on whether individual speaker-listeners' production patterns are linked to the information they attend to in perception. However, for prosodic structure, the production-perception relation has received little attention. This dissertation investigates the hypothesis that individual participants vary in their production and perception of prosodic boundaries, and that the properties they use to signal prosodic contrasts are closely related to the properties used to perceive those contrasts.

In an acoustic study, 32 native speakers read eight sentence pairs in which the type of prosodic boundary (word and Intonational Phrase boundary) differed. Phrase-final and initial temporal modulation, pause duration, and pitch reset at the boundaries were analyzed. Results showed that, as a group, speakers lengthened two phrase-final syllables, shortened the post-boundary syllable, and produced a pause and pitch reset when producing an IP boundary. However, individual speakers differed in both the phonetic features they used and the degree to which they used them to distinguish IP from word boundaries. Speakers differed in the onset and scope of phrase-final lengthening and presence of shortening (resulting in six different patterns), pause duration, and the degree of pitch reset at the IP boundary, including in ways that demonstrated a trading relation between these properties for some individuals and an enhancement relation for others. The results suggest that individuals differ in how they encode prosodic structure and offer insights into the complex mechanism of temporal modulation at IP boundaries.

In an eye-tracking study that tested the perceptual use of these acoustic properties by 19 of these same participants, the productions of a model talker were manipulated to systematically vary the presence and degree of IP boundary cues. Twelve unique combinations of cues, based on the main patterns in the production study, were created from four phrase-final lengthening patterns, two pause durations (presence/absence of a pause), and three pitch reset values. Patterns of fixation on the target boundary image over time showed that, as a group, listeners attended to the information conveyed by pause duration and final lengthening as that information became available, with pause being the most salient cue for IP boundary perception.

A clear pattern did not emerge for pitch reset. Adding to the body of research on weighting of the acoustic properties for IP boundary, these results characterize the time-course of the perceptual use of different combinations of IP boundary-related properties.

To examine the production-perception relation, a series of perceptual models in which each participant's average production values were entered as predictor variables tested whether the production patterns are reflected in the same individuals' perception. The results did not provide statistically significant evidence of a production-perception relation, although a trend in the pause duration models across three different conditions was suggestive of a pattern in which individuals with longer pause durations were faster to fixate on the IP boundary target than those with shorter pause durations. The lack of evidence of a close production-perception relation for individual speaker-listeners is inconsistent with the main hypothesis but is in line with the results of several previous studies that have investigated this relation for segmental properties. Further investigation is needed to determine whether, despite the absence of a strong production-perception relation, specific individuals might nonetheless show the link predicted by some theoretical approaches.

CHAPTER 1

Introduction

Phonetic theories must explain successful communication: how do listeners understand the speech produced by a speaker? Understanding the mechanisms underlying successful communication requires taking into consideration the high variability in the speech signal. Researchers have documented many factors that contribute to variable production, such as regional dialect, anatomical differences between speakers, socio-indexical information about the interlocutor, and segmental context (Johnson et al., 1993; Ohala, 1993; Bent & Holt, 2017). In recent years, there has been growing interest in whether and how the systematic variability that arises from individual differences in production may be linked to individual differences in perception. A paradigmatic shift in theoretical approaches to speech perception in recent decades is the change in perspective towards variation in the input – including variation due to individual speakers, which is viewed not as noise, but as information that may guide speech perception (Goldinger, 1996; Norris et al., 2003; Cutler et al., 2010).

This dissertation investigates whether and how individual differences in the production of prosodic boundaries are related to differences in the perception of prosodic boundaries. Although previous research has reported evidence of substantial individual variation for multiple aspects of prosody (Swerts et al., 1994; Fougeron & Keating, 1997; Byrd et al., 2006; Cole et al., 2010a), systematic investigation of individual differences has mostly focused on the production and/or perception of prominence (Cole et al., 2010b, Cangemi et al., 2015, Grice et al., 2017, Roessig & Mücke, 2019, Roessig et al., 2019), with Roy et al. (2017) being the exception with their study of individual differences in prosodic boundaries. Most studies on perception of prosodic boundaries have investigated how different cues to IP boundaries are weighted, and the non-uniform findings (Lehiste et al., 1976; Scott, 1982; Beach, 1991; Mo & Cole, 2010; Zhang, 2012) serve as further motivation for this study.

The current study investigates the production of acoustic properties of Intonational Phrase (IP) boundaries in American English and its relation to the perception of these cues by the same individuals using eye-tracking. The goal of the study is to delineate inter-speaker differences in the combination and degree of the acoustic properties used to distinguish IP boundaries from word boundaries and to test whether individuals' patterns observed in production are mirrored in their perceptual cue-weighting strategies. By examining whether individuals manifest speaker-specific strategies to signal prosodic boundaries that in turn reflect these individuals' own perception of prosodic structure, the study aims to extend our understanding of the mechanisms of speech production and perception.

This chapter discusses the background and theoretical underpinnings for the study. The first section of the chapter (1.1) discusses research on the prosodic hierarchy of American English. The next section (1.2) reviews the findings of previous studies that have investigated the production and perception of IP boundaries. The following section (1.3) elaborates on the studies that investigated speaker/listener-specific ways of producing and perceiving segmental properties as well as prosodic structure. Finally, the last section (1.4) presents the research questions and hypotheses of the current study and introduces the design of the study.

1.1. Prosodic hierarchy of American English

Prosody can be defined in various ways (Ladd, 2008). The definition adopted here is that it is the linguistic structure above the word level that signals prominence and phrasing by varying phonetic properties of the utterance. Prosody marks the information structure and rhythmic structure of the utterance by giving emphasis to particular words that carry important meaning, and it signals how the utterance is organized by grouping words into bigger chunks (Oller, 1973; Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986; Shattuck-Hufnagel & Turk, 1996; cf. Chodroff & Cole, 2018). While prosodic structure is largely determined by syntactic structure, it is now generally agreed that the prosodic structure and the syntactic structure of the utterance are not always isomorphic (Shattuck-Hufnagel & Turk, 1996; Cole, 2015). A variety of different models of prosodic structure have been suggested (see overview in Shattuck-Hufnagel & Turk, 1996; Wagner & Watson, 2010), and they vary in the number and type of proposed phrases above the word level. While the largest and smallest constituents are less equivocally classified, there is more disagreement in the middle of the hierarchies.

The current study is conducted within the framework of the Autosegmental-Metrical theory of intonational phonology (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Ladd, 2008). Developed within the tradition of generative phonology, the Autosegmental

Metrical (AM) model is a phonological theory of intonational structure. It states that tones are autosegments that are independent of speech segments such as consonants and vowels. Tones are theorized as units of contrast, similar to segment-level distinctive features; they are discrete elements of intonation that can be illustrated as Low (L) and High (H), which can stand alone or concatenate to describe various metrical configurations. One of the goals of the model is to provide a uniform account for describing melodic and rhythmic components of spoken language that exist at the lexical level as well as the phrasal level. Within this model, intermediate phrase (ip) and Intonational Phrase (IP) are two prosodic categories above the level of the word in the prosodic hierarchy. This study investigated the production and perception of the IP boundary.

The IP is the highest phrasal boundary category in the AM model (Beckman et al., 2005). Studies across a variety of languages have shown that the most salient phonetic properties of IP boundaries are: boundary-related lengthening (typically phrase-final lengthening), a silent pause, and a pitch reset (Lehiste et al., 1976; Scott, 1982; Edwards et al., 1991; Wightman et al., 1992; Berkovits, 1993; Ferreira, 1993; de Pijper & Sanderman, 1994; Swerts et al., 1994; Venditti et al., 1996; Fougeron & Keating, 1997; Swerts, 1997; Byrd & Saltzman, 1998; Cho & Keating, 2001; Zvonik & Cummins, 2002; 2003; Byrd et al., 2006; Krivokapić, 2007). For example, the two sentences below differ in the type of boundary (shown hereafter as a pound sign) between the words 'her' and 'Melinda'. The words in (1) are separated by an IP boundary while the words in (2) are separated by a word boundary. The difference in the type of boundary results in different meanings of the two utterances, such that 'Melinda' in (2) is referring to 'her' whereas 'Melinda' in (1) does not necessarily identify 'her' but rather begins a new phrase.

(1) Dad called her. # Melinda and Paul said hello.
IP boundary

(2) Dad called her # Melinda. And Paul said hello.
Word boundary

To illustrate the three primary cues for an IP boundary compared to a word boundary using this example, the word 'her' would be produced with longer duration before the IP boundary in (1) than before the word boundary in (2). A silent pause is likely to occur after 'her' in (1) but not in (2). Lastly, the difference in pitch levels between the words 'her' and 'Melinda' should be greater in (1) than in (2).

1.2. Production and perception of prosodic boundaries

1.2.1. Three primary phonetic properties of IP boundaries in American English

Early studies showed that there is a clear distinction between phrase-final versus phrase-internal segment durations, such that speakers produce final lengthening to mark a strong prosodic boundary and listeners may use it to detect that boundary (Oller, 1973; Lehiste, 1973; Lehiste et al., 1976; Klatt, 1975; Scott, 1982). For example, Oller (1973) and Lehiste (1973) showed that phrase-final segments and syllables in nonsense target words (e.g., /babababab/) in a carrier sentence were significantly longer than non-final segments and syllables. Klatt (1975) found phrase-final lengthening in phrase-final vowels and syllables in read speech.

Using a speech corpus of four speakers of American English reading 35 pairs of syntactically ambiguous (but prosodically disambiguated) sentences, Wightman et al. (1992) examined the durational characteristics of prosodic boundaries. They found a significant correlation between the normalized duration of the rhyme (vowel nucleus and any coda consonant) of the phrase-final syllable and the perceived size of the boundary expressed as break indices of 0 to 6, such that a longer rhyme duration signaled a larger boundary. They also showed that the duration of the vowel nucleus of the last pre-boundary syllable may perceptually distinguish at least four levels of prosodic boundaries, though they pointed out that more boundaries could potentially be distinguished if other phonetic properties were considered.

A large body of studies have investigated boundary-related lengthening in acoustic (Berkovits, 1993, 1994; Cambier-Langeveld, 1997; Shattuck-Hufnagel & Turk, 1998; Turk, 1999; Turk & Shattuck-Hufnagel, 2007) and articulatory (Edwards et al., 1991; Fougeron & Keating, 1997; Byrd & Saltzman, 1998; Byrd, 2000; Byrd et al. 2006; Krivokapić, 2007) domains. These studies examined acoustic and articulatory events at IP boundaries to determine the extent of boundary-adjacent lengthening. A general finding across studies is that the rhyme of the phrase-final syllable manifests phrase-final lengthening consistently and robustly. Wightman et al. (1992) showed that the lengthening is limited to the rhyme that immediately precedes the prosodic boundary. The segments between the foot-initial vowel and the final vowel as well as the foot-initial vowel of the lexically stressed syllable in words with an unstressed word-final syllable did not correlate with perceived boundary strength. However, in that study, segments between the foot-initial vowel and the final vowel before the boundary were not each tested separately. Rather, they were tested for correlation as a single interval, which might have obscured any spreading of the boundary-related lengthening in any of the intervening segments.

Shattuck-Hufnagel & Turk (1998) investigated the domain of phrase-final lengthening in tri-syllabic words with lexical stress on either the first or second syllable and found that the syllables with lexical stress manifest significant lengthening, in addition to the phrase-final syllables. Turk & Shattuck-Hufnagel (2007) expanded the comparison to words with different numbers of syllables, stress patterns, and phonological composition of the phrase-final syllable. They found that, when the syllable with primary lexical stress was located earlier than the phrase-final syllable, the intervening syllable showed less or no final lengthening compared to the phrase-final syllable and the stressed syllable.

As for a rightward, post-boundary effect, there are relatively few studies, and these studies report mixed findings – and individual variation – regarding presence of the effect (Oller, 1973; Wightman et al., 1992; Fougeron & Keating, 1997; Byrd et al., 2006; Cho & Keating, 2009 for English; Hsu & Jun, 1998 for Taiwanese; Katsika, 2016 for Greek). For instance, while Wightman et al. (1992) did not find phrase-initial lengthening, Hsu & Jun (1998) found lengthening of phrase-initial consonants, and Cho & Keating (2009) found lengthening of the phrase-initial consonant and the following vowels. Byrd et al. (2006) and Katsika (2016) also showed lengthening of the phrase-initial consonant.

Within the framework of Articulatory Phonology (Browman & Goldstein, 1992, 1995; Goldstein, Byrd & Saltzman, 2006), prosodic boundaries are modeled as gestures (Byrd & Saltzman, 2003). The prosodic gesture (π-gesture) locally slows the time flow of gestural activation thereby slowing constriction gestures that are co-active with it and, as a consequence, gestures become longer (and also slower and less overlapped) (Byrd & Saltzman, 2003). Under this model, all constriction gestures within the activation of the π-gesture are affected, there is one continuous domain of boundary-adjacent lengthening, which manifests its effects on all articulatory gestures (and by extension segments) that are co-active with the π-gesture, and the scope of lengthening is determined by the temporal extent of the π-gesture. Evidence for a local and continuous domain of lengthening has been found in a number of studies (e.g., Berkovits, 1994 for Hebrew; Cambier-Langeveld, 1997 for Dutch; Byrd et al., 2006 for English; Katsika, 2016 for Greek). There is also evidence that the scope of phrase-final lengthening interacts with prominence (Byrd & Riggs, 2008; Katsika et al., 2014). In American English, Byrd & Riggs (2008) found that one out of three speakers produced lengthening of a stressed syllable preceding the phrase-final syllable. In Greek, Katsika (2016) found an interaction between the location of lexical stress and the onset of phrase-final lengthening. Phrase-final lengthening began later in phrase-final words with final stress than in phrase-final words where stress occurs earlier, indicating that stress shifts the onset of the boundary towards the stressed syllable (for a detailed account of the interaction of boundaries and prominence see Katsika et al., 2014). Their results provide additional support for the π-gesture model by showing continuous boundary-related lengthening that was more robust closer to the boundary.

Turning to pitch, Ladd (1988) tested the hypothesis that a stronger boundary would be associated with a larger declination reset following the boundary. He used phrases in which the same words are joined by two different conjunctions – i.e., “A and B but C” and “A but B and C”, assuming that the words joined by “but” straddle a stronger boundary than the words joined by “and”. He analyzed the pitch top lines (i.e., the pitch peak associated with the three accented syllables; A, B, and C in the example above) of the clauses, along with the pitch end points (i.e., the f0 values at the end of clauses A, B, and C), as well as the durations of the phrase-final segments and the following pause (in order to independently confirm that the sizes of the boundary differed). There was a general downward trend over the course of an utterance in terms of the pitch top lines, such that the f0 peak at the end of B was lower than at the end of A, and f0 peak at the end of C was lower than at the end of B. A consistent pattern of pitch reset emerged when the post-boundary peaks (at the beginning of B and C) were compared across experimental conditions (the two different conjunctions). In all participants’ production of the sentences, the peak was higher – i.e., the reset was larger – at the boundary of the conjunction ‘but’ than at the boundary of the conjunction ‘and’. The lowering of f0 associated with the end of IPs has been documented in other studies of production (van den Berg et al., 1992; de Pijper & Sanderman, 1994; Swerts, 1997; Lin & Fon, 2011; Truckenbrodt & Féry, 2015).

Regarding pauses, Ferreira (1993) showed that the duration of a silent pause reflects the prosodic structure of the utterance rather than the syntactic structure. Findings also show that pause duration is positively correlated with the strength of the prosodic boundary (Horne et al., 1995; Krivokapić, 2007). Other studies have noted the co-occurrence of silent pauses and IP boundaries (Swerts, 1997; Petrone et al., 2017) and individual speaker variation in the pause duration (Fant et al., 2003).

1.2.2. Perception of the three primary phonetic properties of IP boundaries

Perception studies have shown that these properties of prosodic boundaries guide speech processing. Streeter’s (1978) findings for acoustic correlates of major phrase boundaries led her to argue that duration and pitch are primary cues for perceiving major phrase boundaries. In a perception experiment in which English-speaking participants distinguished syntactically ambiguous sentences with or without phrase-final lengthening and/or a pause at a prosodic boundary, Scott (1982) found that listeners use these temporal characteristics to parse syntactically ambiguous sentences. Price et al. (1991) also used syntactically ambiguous sentences to investigate the role of phonetic properties of prosodic boundaries, such as phrase-final lengthening, pause, and boundary tones in disambiguating syntactic ambiguity. Their results showed that listeners reliably disambiguated the target sentences using these prosodic cues.

Using a series of rating experiments, Swerts et al. (1994) examined pitch register (low and high), pitch range, and pitch contour to show that these tonal cues signal prosodic boundaries for listeners. Swerts & Geluyskens (1994) also reported that listeners use the tonal markers and pauses to signal prosodic boundaries in the absence of semantic cues. A number of other studies have investigated silent pauses in relation to prosodic structure (de Pijper & Sanderman, 1994; Sanderman & Collier, 1995 for Dutch; Swerts, 1997; Cho & Hirst, 2006 for Korean; Roy et al., 2017 for English).

Some more recent studies have used techniques that investigate moment-by-moment processing of prosodic information. Studies examining neural correlates of prosodic marking, for example, found that the Closure Positive Shift (CPS) in event-related potentials (ERPs) is associated with prosodic boundaries (Steinhauer et al., 1998; Steinhauer et al., 1999; Steinhauer & Friederici, 2001; Steinhauer, 2003; Männel & Friederici, 2009; Roll & Horne, 2011; Männel et al., 2013; Holzgrefe et al., 2013). Männel & Friederici (2016) showed that CPS was evoked only when some boundary-related cues were present in the auditory stimuli. Holzgrefe-Lang et al. (2016) used a prosodic judgement task in combination with ERP methodology in German to investigate perception of different combinations of two acoustic cues to IP boundary: pitch change at the end of prosodic phrase and final lengthening. They found that boundary perception in the form of CPS was elicited when both cues were present but not when only one of them was present.

Results from eye-tracking studies have shown that listeners rapidly integrate prosodic information for boundary-marking (Lee et al., 2008) and prominence-marking (Weber et al., 2006; Ito & Speer, 2008; Kurumada et al., 2014). For example, Snedeker & Trueswell (2003) examined the time course of listeners' use of prosodic cues in processing ambiguous utterances. Their results suggest that prosodic cues prior to the ambiguous portion of the utterance are available to listeners and influence their initial interpretation of the utterance.

Very few studies have used on-line methods to tease apart the effect of multiple cues for prosodic boundaries that become available at different points in time, and there is no study as of yet that has investigated individual listeners' different sensitivity to those multiple cues and that sensitivity's relation with the same individuals' production.

1.2.3. Relative perceptual importance of the primary cues to IP boundaries

While it is known that many sources of information are available to listeners to detect the presence of major phrase boundaries, including amplitude (Streeter, 1978) and irregular pitch periods (Cole & Shattuck-Hufnagel, 2011), researchers have predominantly focused on attention to three acoustic cues to major prosodic boundaries: final lengthening, pitch reset, and pause duration. However, studies have yielded inconsistent results in terms of the relative weighting of these properties.

The results of multiple studies suggest that phrase-final lengthening is more heavily weighted than pause or pitch reset by English-speaking listeners (Lehiste et al., 1976; Scott, 1982). Streeter (1978) found not only that duration is more important than pitch, but also that the combination of these cues had a greater perceptual effect than the duration cue alone. Other results, though, suggest a greater role of pause (Zhang, 2012) or pitch (Seidl, 2007; Bögels & Torreira, 2015) in the processing of prosodic boundary. Beach (1991) argued that there is a trading relationship between duration and pitch cues in perception of structurally ambiguous utterances, and Ferreira's (1993) results showed an inverse relationship between pause duration and pre-boundary word duration in production, which may be relevant to the way that individuals with different relation between final lengthening and pause duration use these temporal cues to perceive IP boundaries.

There are also mixed results on the weighting of these cues by speakers of languages other than English. For Mandarin, Shen (1992) found that Mandarin-speaking listeners relied more heavily on pauses than on phrase-final lengthening to identify a major prosodic boundary. Similarly, Yang et al. (2014) investigated the relative perceptual weighting of the primary acoustic cues to prosodic boundaries of Mandarin and concluded that listeners found pause to be a more robust cue for IP boundary than phrase-final lengthening or pitch reset. Other studies on Mandarin, though, identified the pitch cue (Zhang, 2012) or the pause (Lin & Fon, 2011; Yang et al., 2014) to be more heavily weighted than the other cues by listeners. For German, Wellman et al. (2012) reported that pause was not a reliable marker for IP boundaries for infant listeners learning German. However, Petrone et al (2017) argued that, while all three cues were reliably used by German adult listeners to perceive a prosodic boundary, final lengthening and f0 were perceived in a gradient manner, whereas pause perception was more categorical and therefore a reliable marker for IP boundaries. Petrone et al. (2017) also cited Peters' (2005) results in which the role of pause was found to be more important than the other cues.

1.3. Individual differences in the production-perception relation

A foundational issue in speech perception research has been to determine how the listener extracts a linguistic message from the input signal despite the complex mappings between the linguistic units and their acoustic realizations. Although contemporary theoretical approaches to speech perception differ in fundamental aspects, some of these approaches including gesturalist theories (Liberman & Mattingly, 1985; Fowler, 1986) and exemplar theories (e.g., Goldinger, 1997; Pierrehumbert, 2002), share a common understanding that successful communication requires parity between the forms of speaking and the forms of listening. What needs to be explained, then, is the nature of this parity. One way to examine this question is to investigate the relation between speaking and listening at the individual level, by examining individual speaker-listener's strategies for the production and perception of multiple acoustic cues for a targeted linguistic property.

Previous studies investigating the relation between production and perception for individual speaker-listeners have focused on segmental properties including /u/-fronting (Kataoka, 2011), vowel-to-vowel coarticulation (Grosvald & Corina, 2012), co-varying cues for stop voicing (Shultz et al., 2012; Schertz et al., 2015; Coetzee et al., 2018) and anticipatory vowel nasalization (Beddor et al., 2018), and their findings are not uniform. For example, of the six studies just cited, only two reported a systematic relation between individuals' production and perception patterns. Beddor et al. (2018) showed that American English listeners' perceptual use of coarticulatory vowel nasalization was reflected in the same individuals' timing of production of nasal coarticulation. Coetzee et al. (2018) showed a weak but significant correlation between perceiving and producing co-varying VOT and f0 information for individual Afrikaans speaker-listeners.

The current work examines how individual speakers employ multiple cues for prosodic boundaries, and whether those cue weights are reflected in their own perceptual biases regarding useful information for prosodic boundaries. Although investigation of this relation is motivated, as suggested above, by theoretical approaches to speech perception, it is also motivated by an interest in the phonetic sources of sound change. In particular, studies of the initiation of perceptually motivated sound changes are grounded in the fundamental assumption that listeners' perceptual strategies are reflected in their own productions (Ohala, 1981; Harrington et al., 2008; Beddor, 2009). For example, for perceptually motivated changes involving coarticulatory variation, Ohala (1981) suggested that listeners may imperfectly identify the source of coarticulation and interpret the coarticulatory property as inherent to the phonetic signal, which arguably leads listeners-turned-speakers to reproduce the misheard signal in their subsequent productions – a "mini" sound change. An alternative to a misperception account recognizes the potential role of individual differences in how listeners assign relative weights to co-varying cues. In this case, the different cue-weights may be reflected in listeners-turned-speakers' productions, which again has the potential to contribute to a shift in the phonetic norm for a speech community (Beddor, 2009; Beddor et al., 2018; Kuang & Cui, 2018). By examining individual strategies for producing and perceiving prosodic boundaries, this dissertation tests the important assumption that listeners manifest their own perceptual biases in their production.

1.3.1. Individual variation observed in production and perception of prosodic boundaries

Numerous studies have investigated how individuals differ in various aspects of realizing segmental distinctions (Dilley & Shattuck-Hufnagel, 1995; Escudero et al., 2009; Kong & Edwards, 2011; 2016; Kong et al., 2012; Shultz et al., 2012; Idemaru et al., 2012; Schertz et al., 2015; 2016; 2019; Kim et al., 2018; Beddor et al., 2018). However, individual differences in encoding and decoding prosodic structure, and boundaries in particular, have received less attention. (The focus here is on prosodic boundaries; for work on individual variation in production and perception of prominence, see e.g., Cole et al., 2010b; Niebuhr et al., 2011; Mücke & Grice, 2014; Roessig & Mücke, 2019, Roessig et al., 2019.)

Previous studies have noted substantial variation across speakers in the production of prosodic boundaries. For instance, Fougeron & Keating (1997) examined linguopalatal contact during /n/ in different prosodic positions in a sentence using electropalatography (EPG). Participants were asked to repeat sentences in which syllables were replaced with a reiteration of the syllable 'no' ([no]). The sentences were arithmetic statements, and therefore different symbols between numbers induced different phrasing of the segmentally identical sentences. Measurements included acoustic duration and linguopalatal contact of each C and V of the three syllables in a numeral (e.g., eighty-nine spoken as [nonono]) placed in Utterance-initial, Utterance-medial, IP-initial, IP-medial, IP-final, PP-initial, PP-medial, PP-final, Word-initial, Word-medial, and Word-final locations.

Overall, Fougeron & Keating found that speakers distinguish prosodic boundaries in articulation and in the resulting acoustic signal. For example, they found greater linguopalatal contact for syllable-initial [n]s at larger prosodic phrases than for those at smaller prosodic phrases and less contact for the vowel phrase-finally at larger phrases than at smaller boundaries (both indicating more extreme articulation). However, the three speakers varied in how they distinguished the prosodic units by the degree of linguopalatal contact. Speaker 1 produced significantly less contact for the vowel IP-finally than PP-finally, and less contact PP-finally than Word-finally. In contrast, Speaker 2 and Speaker 3 did not differ in degree of contact for the IP-final vowel and PP-final vowel. Participants' productions also differed in the acoustic durations of the final vowel in each domain: Speaker 3 distinguished IP-final and PP-final vowel durations by lengthening IP-final vowel duration, whereas Speakers 1 and 2 did not show lengthening differences between IP-final and PP-final vowels. A cross-linguistic follow-up study (Keating et al., 2003) reported similar individual speaker variation in three other languages (French, Korean, and Taiwanese Mandarin).

Byrd et al. (2006) used electromagnetic articulography (EMA) to analyze both temporal and spatial dimensions of boundary-adjacent articulatory movements and examined the temporal scope of prosodic boundary effects. Both articulatory and acoustic data revealed variation across four speakers. For example, for three speakers, pause durations ranged between 200 and 970ms, while for the fourth speaker, most durations were shorter than 200ms. Interestingly, for this speaker, pre-boundary lengthening extended further leftwards than for the other speakers.

Cole et al. (2010a), using the spontaneous conversational speech from the Buckeye corpus, observed inter-speaker variability and (indirect evidence of) inter-listener variability in their study of prosody perception in American English. They developed a new method for prosodic annotation, Rapid Prosody Transcription (RPT), in which untrained listeners transcribe prosodic aspects of speech in real time. In order to assess inter-speaker and inter-transcriber variability, Cole et al. calculated probabilistic Boundary scores (B-scores) for each word, which was the proportion of transcribers from the total group who marked a prosodic boundary following each word in an utterance. Therefore, B-scores are between 0 and 1, with 0 meaning that no transcriber perceived a boundary following that word, and 1 meaning that all transcribers perceived a boundary. The results showed variability in both the production and annotation data. For example, speakers varied in the mean interval between boundaries and between prominent words, as judged by listeners, such that some speakers produced prosodic boundaries and prominent words at comparable intervals, while for some other speakers the mean interval between boundaries was longer than the mean interval between prominent words, or vice versa. The distribution of B-scores was taken as an indirect measure of speaker variability in boundary production. Listeners were another source of variability in the B-scores: for instance, the grand average of the mean interval between boundary marking across all listeners was 8.2 words, but the mean interval per listener was as short as 4.91 words and as long as 15.5 words.

Roy et al. (2017) observed individual differences in perception in their study of untrained American English listeners' annotations for boundary and prominence marking of an excerpt from the Buckeye corpus. They also conducted an acoustic analysis of the auditory stimuli to evaluate the influence of the presence or absence of prosodic cues for boundary or prominence on inter-rater agreement. The results revealed substantial individual differences across the annotators as well as uniformity among them; the biggest distinction was between those annotators who relied solely on durational cues and those who used one or more cues in addition to the durational cues for boundary marking.

Overall, these studies suggest that individual variation in signaling and perceiving prosodic contrasts is pervasive. However, while empirical findings suggest high variability across individuals in their production (and, to a lesser extent, perception) of various prosodic features, it is the object of investigation in very few studies.

There is only one study that examines how individual speakers' production of prominence is related to (a separate group of) individual listeners' perception of prominence. Cangemi et al. (2015) investigated individual production and perception of intonational contrasts. They hypothesized that some speakers' productions of contrast may be more intelligible to some listeners than others. In their production experiment, native speakers of German were asked to read aloud answers to questions in which the target word appears in a variety of prosodic conditions, thereby eliciting four different focus structures. Both acoustic and articulatory data were collected during the experiment. A separate group of listeners participated in the perception experiment. They listened to utterances from the production study and were asked to match the sentence heard to one of the four questions corresponding to the focus structure conditions, in that way indicating which focus condition they perceived.

In Cangemi et al. (2015), the results of the analysis of the productions of five speakers showed that individual speakers differed in both the number of cues used to mark a particular focus category and in the partitioning of a given cue (i.e., whether the cue is employed to mark two or more focus categories). In the perception task, 20 participants listened to the test sentences produced by the five speakers and matched each utterance to the appropriate question. The results of the analysis of the perception data showed a general trend in which listeners were more accurate in identifying the productions of speakers who were better encoders of focus (measured as the number of cues used to distinguish the three categories). They also found that there were individual differences among the listeners as well: some listeners were in general better than others at identifying focus. Their results overall provide some initial insight into the complex relationship between cue-weighting processes in production and perception of prosodic prominence and reinforce the importance of accounting for the role of individuals in both production and perception.

Overall, very little research has examined individual differences in the perceptual weighting of prosodic cues to major prosodic boundaries. No study has investigated individual differences in the moment-by-moment processing of prosodic boundaries. Moreover, the precise relationship between production and perception of the cues to prosodic boundaries remains unclear, due to the small number of studies that have been conducted to date. It is thus important to investigate the relationship between production and perception of prosodic boundaries, and the role of individual variation in producing and perceiving the boundary cues. The current study undertakes this investigation and assesses whether a tight connection between the production and perception of the prosodic boundaries may be observed for individual language users.

1.4. The current study

The previous sections highlight the recent interest in individual differences in speech production and perception, including interest in possibly linking individual production patterns with individual perception patterns. This interest is due in part to the changing perspective towards individual (and other) variation: as information rather than as noise. However, for the prosodic boundaries, relatively little is known about whether and how the individual variation observed in production may be related to those same individuals' perceptual weighting of that information. The current work is based on the understanding that individual speakers differ systematically from each other in how they convey prosodic structure and focuses on how these individual speaker differences are manifested in signaling prosodic boundaries of American English. Thus, the main goal of the study is to investigate whether individuals' patterns for producing information about prosodic boundaries are reflected in their perceptual use of those sources of information.

An acoustic production experiment and an eye-tracking perception experiment were conducted to investigate (1) how speakers of American English produce the different primary cues marking the phrasal boundary (IP boundaries), and (2) whether these individual patterns are mirrored in these same individuals’ attention to these cues for phrasal organization.

The remainder of the dissertation is organized as follows: Chapter 2 presents the methods, results, and discussion of the production experiment. Chapter 3 presents the methods (except the construction of the utterances which is discussed in the method section in Chapter 2), results, and discussion of the perception experiment. Chapter 4 provides the general discussion of the findings from the two experiments in relation to the hypotheses of the study, with theoretical and practical implications. The conclusion of the study is presented in Chapter 5.

CHAPTER 2

Individual Differences in the Production of Prosodic Boundaries

The experiment reported in this chapter investigates the acoustic correlates of Intonational Phrase (IP) and word boundaries produced by speakers of American English. The main goal of the experiment is to delineate the nature and extent of individual speaker differences in the realization of these boundaries. Based on previous research, it is predicted that speakers will reliably produce acoustic distinctions between IP and word boundaries, but that there will be substantial inter-speaker differences in these acoustic realizations.

2.1. Methods

2.1.1. Stimuli

Thirty-two pairs of sentences, each pair contrasting word vs. IP boundary, were constructed to test how individual speakers vary in their production of three acoustic characteristics of IP boundaries – boundary-related lengthening, pause duration, and pitch reset. Table 2.1 summarizes the conditions and target words.

Prosodic boundary	First target word (TG1)	Second target word (TG2)
IP	maMIma	Melinda
		Belinda
	naNIna	Navarro
		Delilah
Word	maMIma	Melinda
		Belinda
	naNIna	Navarro
		Delilah

Table 2.1. Summary of the experimental conditions. Each boundary (2) x TG1 (2) x TG2 (2) condition occurred in four context sentence types, for a total of 32 sentence pairs.

The target sentences varied in the type of prosodic boundary between the two words in each target word sequence (IP boundary vs. word boundary), which resulted in different meanings of the utterances. Two different first target words (TG1; presented as ‘maMIma’ and ‘naNIna’) were neologisms that were introduced to participants as personal names. Neologisms were used in order to maximize control over all aspects of the production of TG1. Two different consonants were used in TG1 in order to diversify the segmental context (i.e., different places of articulation). To provide more thematic and articulatory diversity in the experimental sentences, two different second target words (TG2) were paired with each TG1 (e.g., ‘maMIma Melinda’ and ‘maMIma Belinda’). The onset consonant of a TG2 always matched the consonant of the corresponding TG1 in place of articulation (bilabial-bilabial or alveolar-alveolar). All four TG2 had the same iambic lexical stress pattern. As a result, eight uniquely different target word sequences were created (2 boundary types x 2 TG1 x 2 TG2). The target sequences were placed within carrier utterances and shown to participants in a randomized order within each repetition. The stimuli for this study were a subset of the stimuli collected as part of a larger study, and no filler items recorded. They were repeated 9 times by 32 participants, and therefore a total of 2,304 utterances were recorded (8 target sequences x 9 repetitions x 32 participants). Examples of the stimuli are presented in Table 2.2. The first pair of sentences (marked with C) provides the context, while the second pair of sentences (marked with T) includes a target word sequence. The pound sign denotes the boundary (word or IP) and was not shown to participants, whereas the boldface denoting words with a contrastive focus and the lower-upper casing in TG1s were shown to participants.

To test the effects of prosodic boundary, the same target word sequence straddles two different prosodic boundaries: IP boundary vs. word boundary. One word in each sentence in the target sentence pair received contrastive focus, correcting the corresponding words in each of the context sentence pair. Because the focused words are located one syllable away from TG1 and TG2, the target word sequence is not pitch-accented.

(1)	TG1 = ‘naNIna’	# = IP boundary
C: The agent called naNIna. # Navarro and Parker bought the painting.		
T: No, the **painter** called naNIna. # Navarro and **Damon** bought the painting.		

(2)	TG1 = ‘naNIna’	# = word boundary
C: The agent called naNIna # Navarro. And Parker bought the painting.		
T: No, the **painter** called naNIna # Navarro. And **Damon** bought the painting.		

(3)	TG1 = ‘maMIma’	# = IP boundary
C: The paramedic called maMIma. # Melinda and Peter said no one got hurt.		
T: No, the **police** called maMIma. # Melinda and **Danny** said no one got hurt.		

(4)	TG1 = ‘maMIma’	# = word boundary
C: The paramedic called maMIma # Melinda. And Peter said no one got hurt.		
T: No, the **police** called maMIma # Melinda. And **Peter** said no one got hurt.		

(5)	TG1 = ‘naNIna’	# = IP boundary
C: The rancher called naNIna. # Delilah and Paige asked about the apples.		
T: No, the **farmer** called naNIna. # Delilah and **David** asked about the apples.		

(6)	TG1 = ‘naNIna’	# = word boundary
C: The rancher called naNIna # Delilah. And Paige asked about the apples.		
T: No, the **farmer** called naNIna # Delilah. And **David** asked about the apples.		

(7)	TG1 = ‘maMIma’	# = IP boundary
C: The king called maMIma. # Belinda and Paul thought that was rude.		
T: No, the **queen** called maMIma. # Belinda and **Daisy** thought that was rude.		

(8)	TG1 = ‘maMIma’	# = word boundary
C: The king called maMIma # Belinda. And Paul thought that was rude.		
T: No, the **queen** called maMIma # Belinda. And **Daisy** thought that was rude.		

Table 2.2. Examples of the stimuli when TG1 is either ‘naNIna’ or ‘maMIma’.

2.1.2. Participants

Thirty-two participants (20 female) were recruited from the University of Michigan campus in Ann Arbor, Michigan. All were native speakers of American English and spent most of their lives in Michigan (i.e., spent less than one year outside of Michigan). None reported hearing, visual, or speech impairment. They received monetary compensation for their participation in an approximately 50-minute task.

2.1.3. Procedures

Prior to the experiment, participants were given instructions for the reading task and sample stimuli containing the context and target sentence pairs containing ‘maMIma # Melinda’ as the target word sequence. They were asked to pay attention to the difference in the meaning of the sentences depending on the different grouping of the two words, and to the boldfaced words being emphasized due to the corrective focus. Participants also listened, over headphones (AKG K240 MKII), to a short (about 50 seconds long) audio file in which the personal names used in the target sentences were introduced by three different female voices in a carrier phrase: “Hi, my name is Melinda” (with the underlined word replaced with different names). The audio file was simultaneously played with the slideshow of the personal names. Upon request from a participant, the audio file and slideshow were played one more time. The purpose of the introduction was to elicit the same pronunciations of the names from all participants.

After the name introduction and prior to the main experiment, there was a short practice session in which four dialogues from the stimuli were read by participants. The purpose of the practice session was to familiarize participants with the structure of the target sentences. In both practice and main experimental sessions, each dialogue consisting of one context sentence pair and the corresponding target sentence pair was presented on a 15-inch computer monitor in large-sized (20 point) font. The context and target sentence pairs were always shown on separate lines. The stimuli were never read aloud to participants by the experimenter. Participants were asked to silently read the context sentence, and then say the target sentence aloud, speaking into a professional microphone. After they finished speaking, the experimenter proceeded to the next dialogue using a remote clicker.

Participants took a 5-minute break in the middle of the experiment. They could take additional 5-minute breaks whenever they needed it, though typically they did not do so. After the experiment, all participants completed an exit survey that contained a series of questions about their linguistic background (see Appendix B).

2.1.4. Acoustic analysis

Three types of acoustic measures were taken for the targeted region of each sentence production to assess how speakers differentiate word and IP boundaries using these measures. An example of the measurements is shown in Figure 2.1.

(1) *Syllable duration*: The durations of the three syllables of TG1, S1, S2, and S3 (e.g., 'maMIma' in Figure 2.1), were measured using Praat software (Boersma et al., 2020) to evaluate the extent of phrase-final lengthening in the pre-boundary direction. The duration of the first syllable of TG2, S4 (e.g., first syllable of 'Melinda', in Figure 2.1), was measured to examine whether it is subject to a boundary effect.

(2) *Pause duration*: Pause duration was measured as the silent interval between TG1 and TG2 in the IP boundary condition. No utterance in the word boundary condition had a pause between TG1 and TG2 and therefore pause duration in these utterances was not measured. Any creaky portion of the word-final vowel of TG1 was not included in pause duration but was included in the vowel duration.

(3) *Pitch reset*: Pitch reset across the boundary between the two target words was measured as the difference between f0 maximum taken from the final syllable of TG1 and f0 maximum taken from the first syllable of TG2. The f0 measurements (in Hz) were first automatically taken using a Praat script, and then manually checked by the experimenter token-by-token to ensure that the measures best represented the f0 maxima during the pre-boundary and post-boundary syllables. Any token in which a correct f0 maximum in either pre-boundary or post-boundary syllable could not be found was excluded from the subsequent statistical analysis for this measure (358 tokens in total).

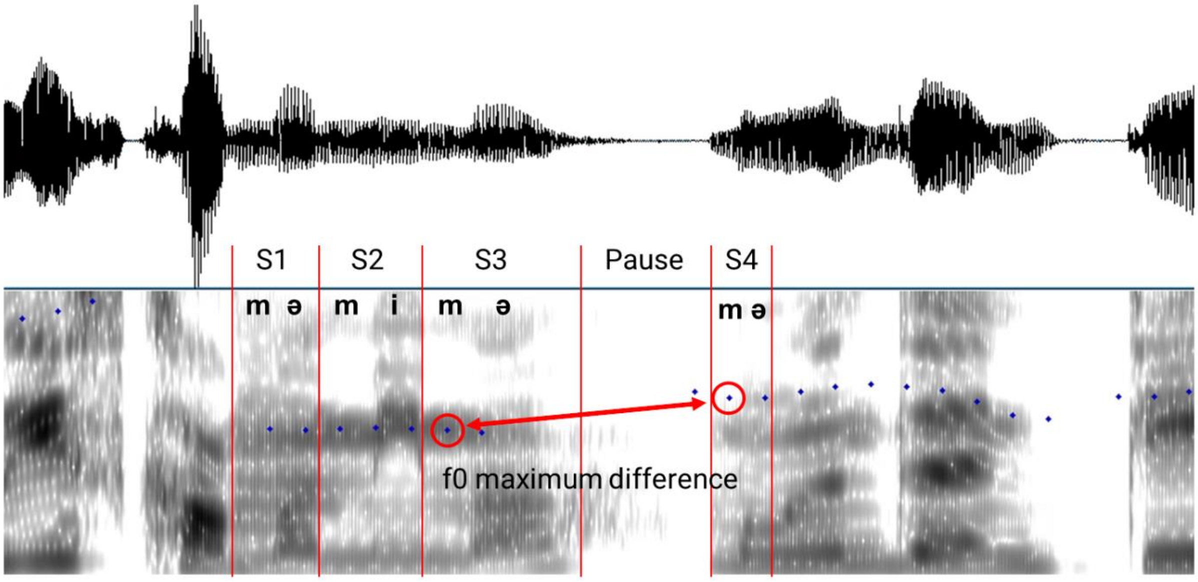


Figure 2.1. Example of the three acoustic measures (syllable durations, pause duration, and f0 maximum difference) from the target word sequence "maMIma # Melinda" produced in the IP boundary condition by participant 9.

During the analysis, all 2,304 utterances produced by the participants were checked for disfluency and placing incorrect boundary by mistake. As a result, 116 utterances (66 'maMIma' and 50 'naNIna' utterances) were excluded from the subsequent statistical analysis. The average number of utterances per participant included in the analysis is 68.38 out of 72 productions per participant.

2.1.5. Statistical analysis

A Linear Mixed-Effects model tested each acoustic measure for the effect of boundary type across speakers, except for pauses since pauses were only measured in the IP boundary condition. The base model included SPEAKER as a random effect, and BOUNDARY as a fixed effect. The base model (Model 1) was then compared to Model 2 using an F test. Model 2 included TG1 TYPE as another fixed effect. (There was no prediction for the effect of TG1 TYPE because the purpose of the different types of TG1 was to introduce variation in the test sentences, given the lack of filler items in the stimuli.) Model 3 included an interaction term between the two fixed effects, and it was then compared to Model 2. The following table summarizes the LM model used for each acoustic measure, in the form of R codes.

Acoustic measure	LM model structure
Syllable 1 (S1)	S1 ~ BOUNDARY + (1| SPEAKER)
Syllable 2 (S2)	S2 ~ BOUNDARY * TG1_TYPE + (1| SPEAKER)
Syllable 3 (S3)	S3 ~ BOUNDARY + TG1_TYPE + (1| SPEAKER)
Syllable 4 (S4)	S4 ~ BOUNDARY + TG1_TYPE + (1| SPEAKER)
F0 maximum difference (Hz)	f0_max_diff ~ BOUNDARY + (1| SPEAKER)

Table 2.3. The model structure for each acoustic measure.

When an interaction between the two fixed effects was significant ($p< .05$), simple regression models on subsets of the data were used to unpack the interaction effect. For example, when the interaction between BOUNDARY and TG1 TYPE is significant, the effect of BOUNDARY was separately tested using a subset of data with 'maMIma' as the TG1 and using the remaining data with 'naNIna' as the TG1. Similarly, the effect of TG1 TYPE was tested using data split by the type of BOUNDARY.

Simple regression models tested whether each acoustic measure, except for pause, was used by individual speakers to distinguish IP boundaries from word boundaries. The base model included BOUNDARY as the only fixed effect, such that the R code was written as "(acoustic measure as dependent variable) ~ BOUNDARY".

The multiple comparisons conducted in the individual analysis may raise concern due to the increased risk of Type I errors (i.e., rejecting true null hypotheses). However, it needs to be pointed out that the widely used adjustment methods such as the Bonferroni adjustment may pose a concern for the increased risk of Type II errors (i.e., failure to reject false null hypotheses) (e.g., Feise, 2002). Therefore, it was decided to not apply the adjustments, with the understanding that this might increase the risk of Type I error.

2.2. Hypotheses and predictions

The hypotheses for the boundary-related effects are as follows: Across speakers, all three measures will be systematically modulated depending on the type of the prosodic boundary. Specifically, the duration of the word-final syllable (S3) of TG1 will be longer in the IP boundary condition than in the word boundary condition. There will be pauses (robust interval of silence) between TG1 and TG2 in the IP boundary condition. Lastly, the f0 difference will be greater across IP boundaries than word boundaries.

Given that previous studies of boundary-related effects found inconsistent results for the durations of the syllables preceding and following S3, it is difficult to make predictions about the scope of phrase-final lengthening. Possible outcomes for the duration of the first syllable (S1) of TG1 are lack of boundary effect or shortening (Krivokapić, 2007; Katsika, 2016). The duration of the second syllable (S2) of TG1 may lengthen or may not undergo temporal modulation (Shattuck-Hufnagel & Turk, 1998, Turk & Shattuck-Hufnagel, 2000; Byrd & Riggs, 2008; Katsika 2016). The duration of the post-boundary syllable (S4) may show lengthening (Fougeron & Keating, 1997; Byrd et al. 2006; Katsika, 2016), shortening (Krivokapić, 2007), or no effect of boundary type (Wightman et al., 1992).

For individual speakers, the hypothesis for the boundary-related effects is that individuals will produce the examined properties to different degrees and in different combinations when distinguishing IP boundaries from word boundaries.

2.3. Results of the production experiment

This section presents the results of the statistical analyses of the boundary-related acoustic effects across the 32 participants (2.3.1.) and individually (2.3.2).

2.3.1. Across all participants

The output of all Linear Mixed-Effects model analyses can be found in Appendix C.1. The analyses showed that, across 32 participants, S1 duration did not significantly differ depending on the type of boundary (p=.194; Figure 2.2a), while the durations of S2, S3, S4 showed a significant effect of BOUNDARY. S2 and S3 were lengthened in the IP boundary condition compared to the word boundary condition (p<.001 for both; Figure 2.2b and 2.2c), whereas S4 was significantly shorter in the IP boundary condition (p<.001; Figure 2.2d).

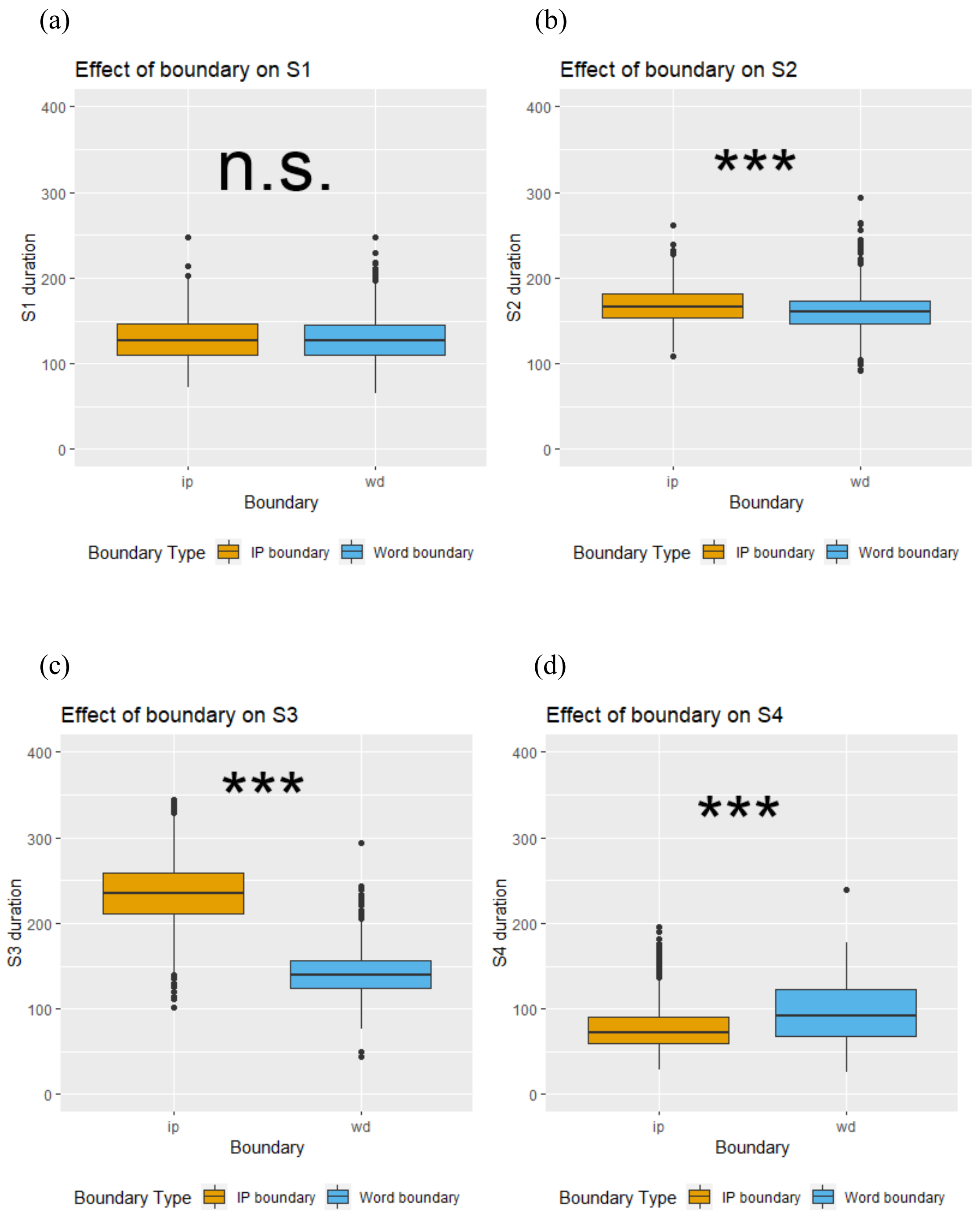
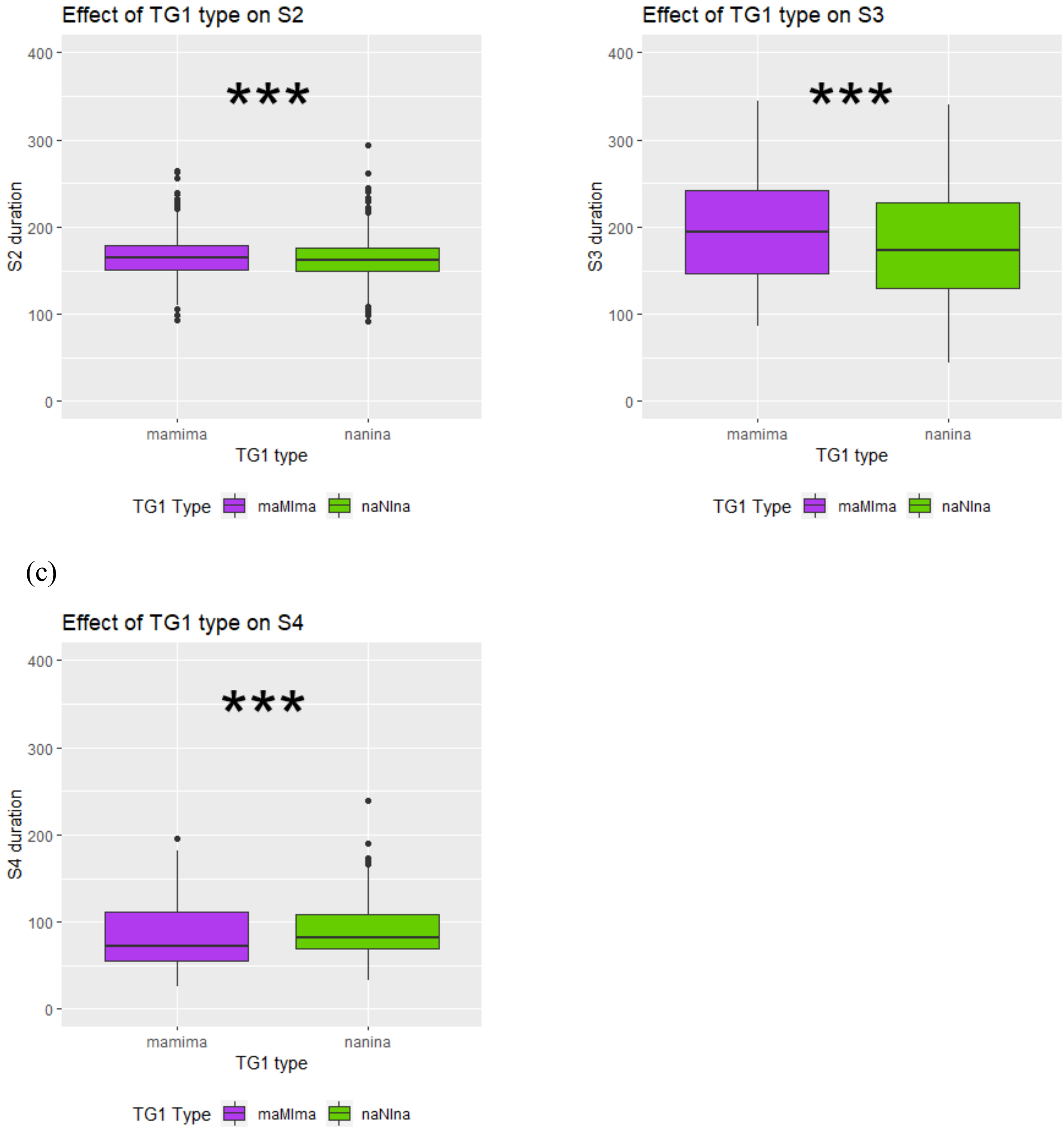


Figure 2.2. Syllable durations of TG1 (S1, S2, S3) and TG2 (S4) by boundary type. (***: p <.001; n.s.: p > .05)

TG1 TYPE was not included in the model as a fixed effect for S1 because it did not significantly improve the model fit (based on the Chi-square Goodness of Fit test (p = .09). It was, though, included in the models for the other syllable duration measures. Those results showed that the effect of TG1 TYPE was significant for S2, S3 and S4 durations. For S2 and

S3, the syllable durations were longer when the pre-boundary target word was ‘maMIma’ compared to ‘naNIna’ (p<.001 for both; Figure 2.3a and 2.3b), whereas S4 duration was shorter when the pre-boundary word was ‘maMIma’ than when it was ‘naNIna’ (p<.001; Figure 2.3c).



(a)

Effect of TG1 type on S2

(b)

Effect of TG1 type on S3

(c)

Effect of TG1 type on S4

Figure 2.3. Syllable durations of TG1 (S2, S3) and TG2 (S4) by TG1 type. (***: p <.001)

Including the interaction between BOUNDARY and TG1 TYPE significantly improved the model fit for S2 (p < .001) and S4 (p=.02), but not for S1 (p=.26) and S3 (p=.97). Unpacking the interaction for S2 duration, the boundary effect was present in both 'maMIma' and 'naNIna' target words (both p<0.001; Figure 2.4a), with S2 duration being significantly longer in the IP boundary than in the word boundary condition. However, the TG1 TYPE effect was present only in the word boundary condition (p<0.001; Figure 2.4b right panel) but not in the IP boundary condition (p=.9; Figure 2.4b left panel), such that, at the word boundary, S2 duration was significantly longer when TG1 was 'maMIma' than when it was 'naNIna'.

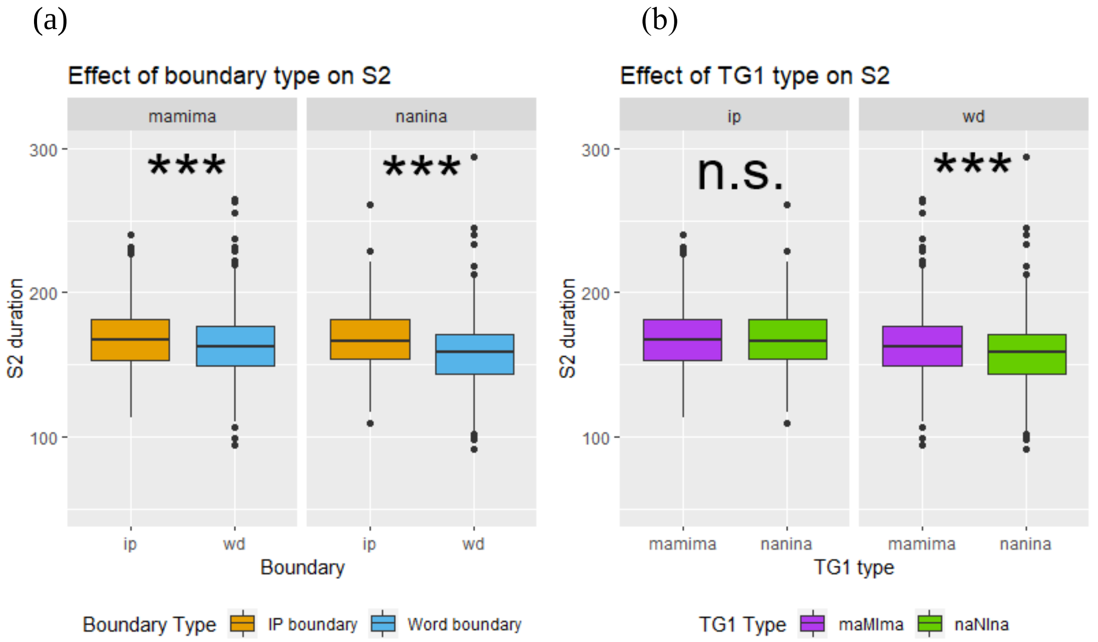


Figure 2.4. Effect of (a) boundary type on S2 duration and (b) effect of TG1 type on S2 duration. (***: p <.001; n.s.: p > .05)

For S4, the effect of BOUNDARY was significant for both 'maMIma' and 'naNIna' (p < .001; Figure 2.5a), while the effect of TG1 TYPE was significant only when the boundary was an IP boundary (p < .001; Figure 2.5b left panel), such that S4 was shorter when TG1 was 'maMIma' than when it was 'naNIna' but there was no significant difference between S4 duration across the word boundary (p = .08; Figure 2.5b right panel).

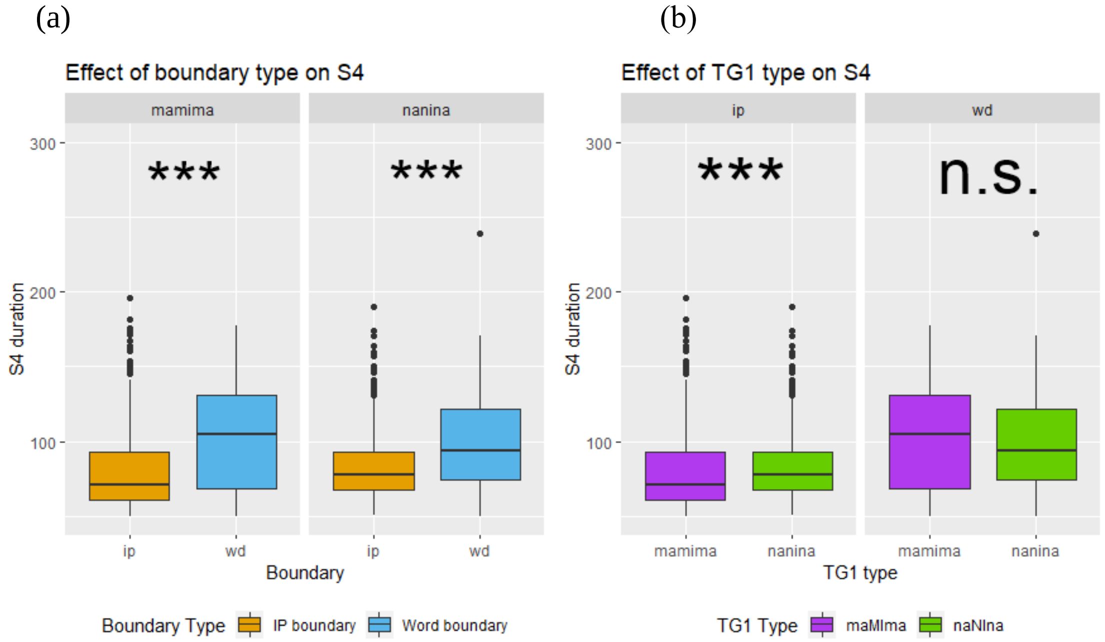


Figure 2.5. Effect of (a) boundary type on S4 duration and (b) effect of TG1 type on S4 duration. (***: p <.001; n.s.: p > .05)

Turning now to the effect of boundary on pitch reset, the analysis of the LM model for Δf0 showed that, across speakers, there was larger pitch reset across the IP boundary than across the word boundary (p<.001; Figure 2.6).

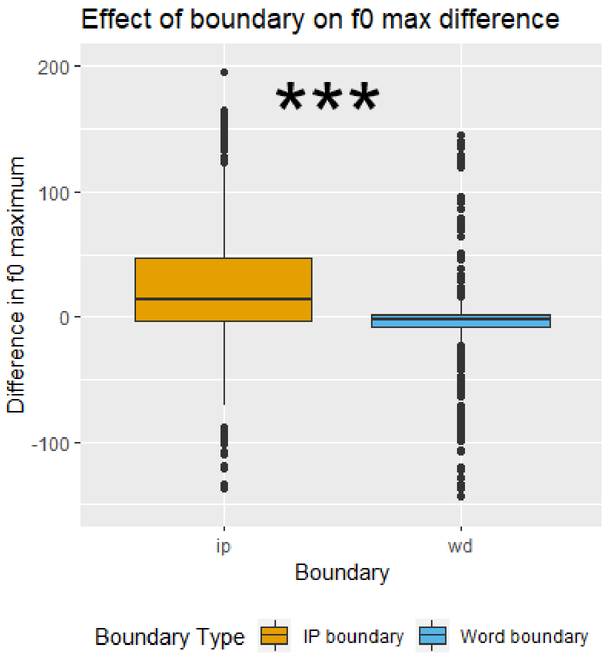


Figure 2.6. Three f0 maximum difference measures by boundary type. (***: p <.001)

2.3.2. *Individual speakers*

The output of the Linear Regressions performed for each participant revealed substantial differences among individual speakers. Both BOUNDARY and TG1 TYPE were included in all linear models tested for each acoustic measure for all participants. The output of all model analyses can be found in Appendix C.2.

2.3.2.1. Syllable durations

The modeled results of the boundary effect on the four syllable durations for 32 individual speakers are shown in Figure 2.7 through 2.10. Twenty-six out of 32 participants did not distinguish IP and Word boundaries on the basis of S1 duration (i.e., the duration of the first syllable of the trisyllabic pre-boundary target word TG1). Four participants had shorter S1 duration before IP than word boundary, while two participants had lengthened S1 before IP than word boundary. The two participants who lengthened S1 also had longer S2 and S3 before IP than before word boundary. One out of the four participants who shortened S1 had longer S2 and S3 before IP than word boundary, whereas the remaining three participants only lengthened S3, but not S2, before IP boundary.

Two participants produced shorter S2 before IP than word boundary, while 18 participants produced longer S2 before IP than word boundary. The remaining 12 participants did not significantly differentiate S2 duration depending on the type of prosodic boundary. All participants lengthened S3 before IP relative to word boundary.

As for S4 duration (i.e., the duration of the first syllable of the post-boundary target word TG2), 23 participants had a shorter S4 after the IP boundary than the word boundary. For the remaining nine participants, there was no effect of prosodic boundary on S4 duration.

The results of the boundary effect on syllable duration for individual speakers are summarized in Figure 2.17 in the discussion section (2.4.1).

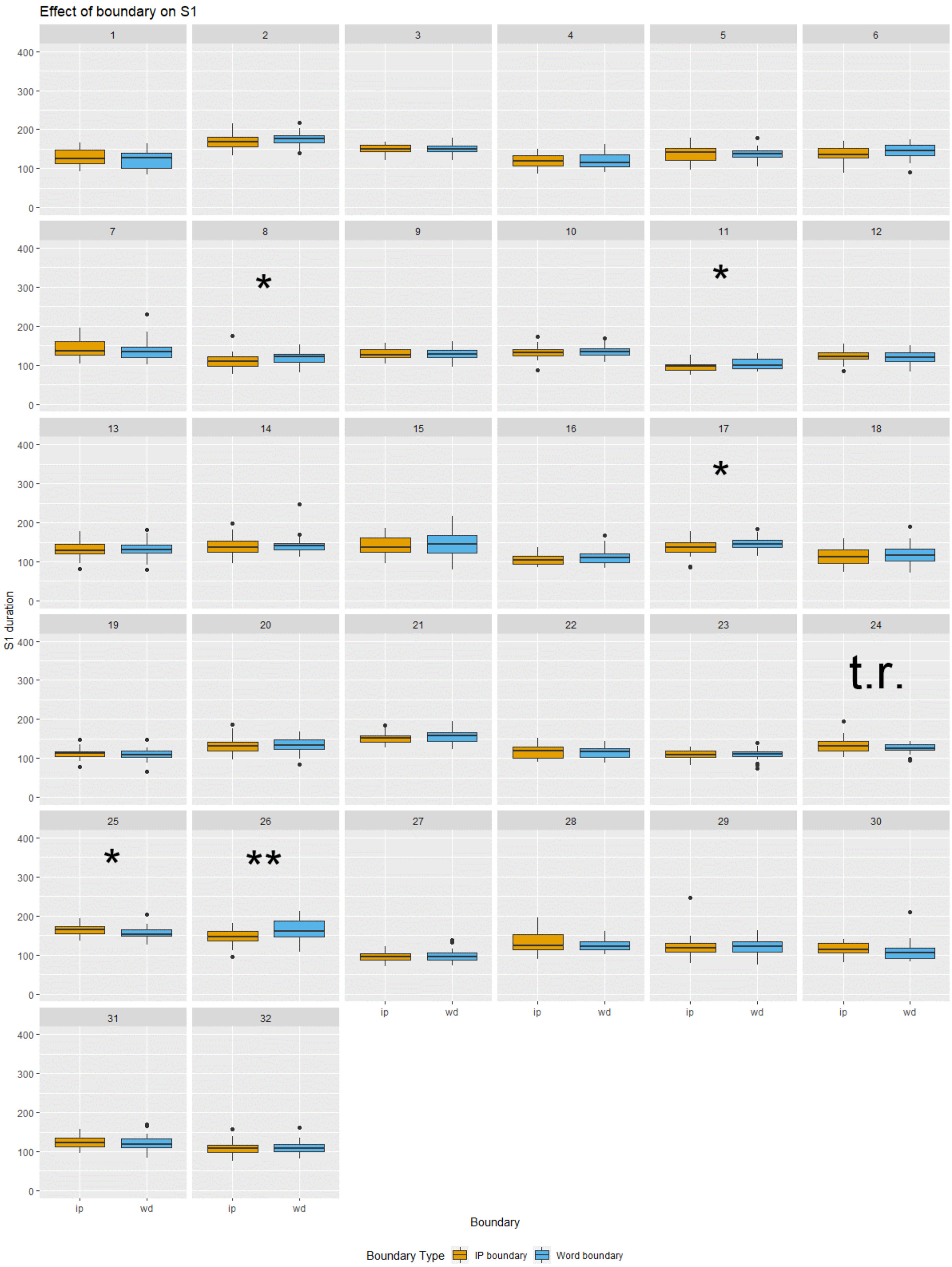


Figure 2.7. Effect of boundary on S1 duration for 32 individual participants. (**: p <.01; *: p < .05; tr: p < .06)

Figure 2.8. Effect of boundary on S2 duration for 32 individual participants. (***: p < .001; **: p <.01; *: p < .05; tr: p < .06)

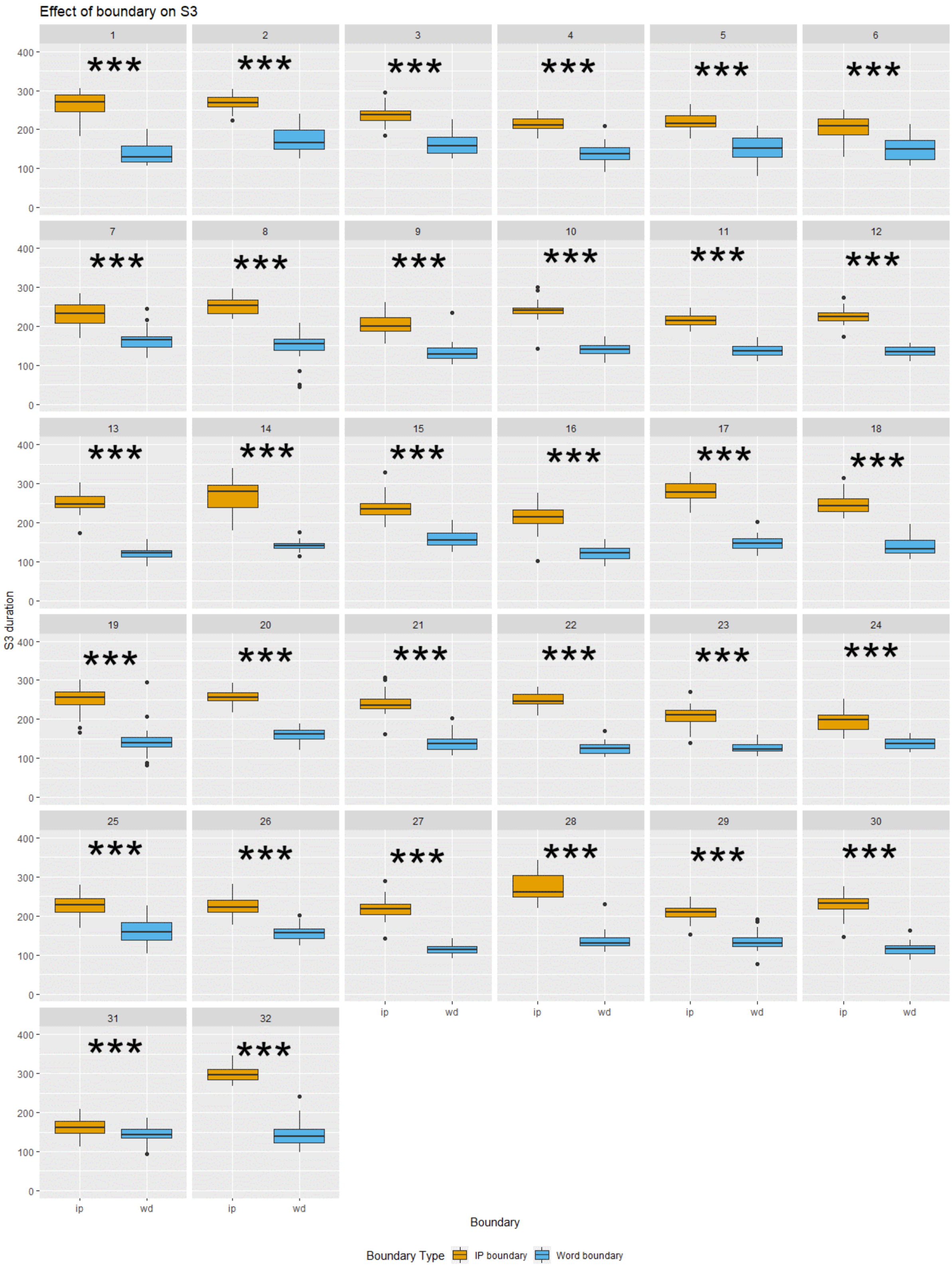


Figure 2.9. Effect of boundary on S3 duration for 32 individual participants. (***: p < .001)

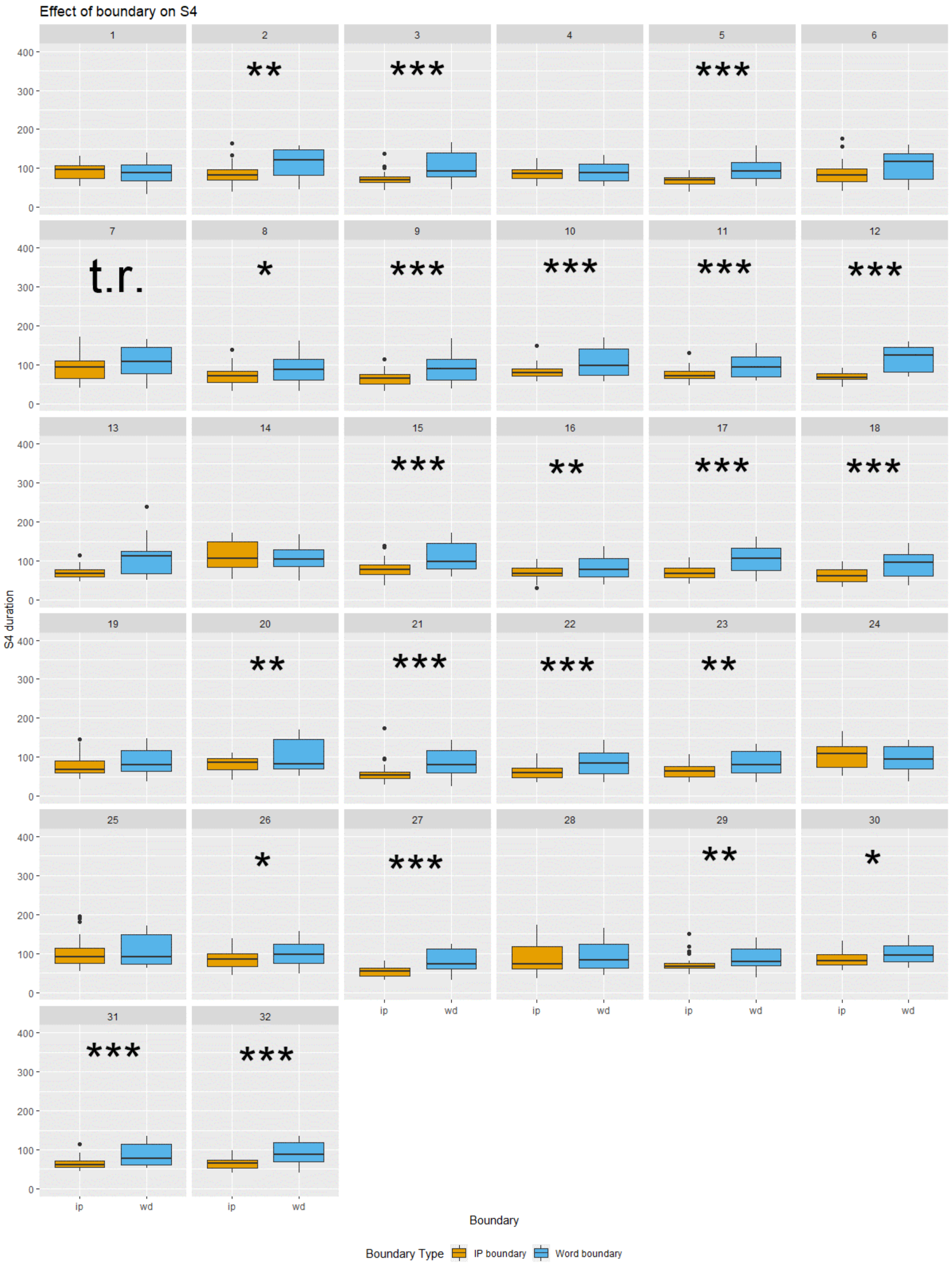


Figure 2.10. Effect of boundary on S4 duration for 32 individual participants. (***: p < .001; **: p <.01; *: p < .05; tr: p < .06)

The results for the effect of TG1 TYPE on four syllable durations are given in Appendix D, and the results for the effect of TG1 TYPE on f0 maximum difference are given in Appendix

E. TG1 TYPE was included in the Linear Mixed-Effects models as a fixed effect based on the model comparison. Since there was no reason to expect any differences based on TG1 (i.e., whether speakers produce 'naNIna' or 'maMIma' differently), no hypothesis about the effect was established before the analysis. Seven participants produced significantly shorter S1 duration when the pre-boundary target word was 'naNIna' than when it was 'maMIma'. Conversely, eight participants produced longer S1 when the target word was 'naNIna'. The remaining 17 participants showed no effect of TG1 type on S1 duration. The pattern for S2 duration was as variable as S1 duration: 10 participants shortened S2 duration when the target word was 'naNIna', while six participants lengthened S2. The remaining six participants did not distinguish S2 duration depending on the target word. As for S3 duration, 25 participants produced significantly shorter S3 when the target word was 'naNIna' than when it was 'maMIma'. Seven participants did not show significant effect of TG1 TYPE. For S4, productions of 11 participants showed a lengthening effect when the target word was 'naNIna', whereas those of only one participant showed a shortening effect in the same condition.

2.3.2.2. Pitch reset

Pitch reset was measured by subtracting the f0 maximum in the last syllable of the pre-boundary target word (S3) from f0 maximum of the first post-boundary syllable (S4). A robust positive value of Δf0 is expected across IP boundaries, suggesting reset, whereas no change in f0 is expected across word boundaries. For each speaker, the effect of boundary type was tested in a Simple Linear Regression with BOUNDARY as the independent variable.

The results for each speaker are shown in Figure 2.11. Table 2.4 summarizes the individual participants' differences between the average f0 maximum difference in word boundary and the f0 maximum difference in IP boundary. A large positive value indicates that the participant consistently produced a pitch reset across IP boundary: for example, for P02, 35.4 Hz in difference of reset indicates a subtraction of -3.9Hz (the average difference between f0 maxima in S3 and S4 across word boundary) from 31.6Hz (the average difference between f0 maxima in S3 and S4 across IP boundary). The output of the statistical model can be found in Appendix C (section 2.2). As expected, a majority of participants (20 out of 32) produced significantly greater Δf0 across the IP boundary than across the word boundary. Eleven participants did not show a significant difference in Δf0 depending on the boundary type. One speaker (P32) produced a small positive reset (average 2.31Hz) at the word boundary and a small negative reset (average -9.29Hz) at the IP boundary, and the difference between these differences was significant. To sum up, 11 out of 32 participants did not distinguish IP boundaries from word boundaries using positive pitch reset, while 21 participants did.

Participant	Difference of reset (Hz)	LR Results
P01	11.3	p=.058 (t.r.)
P02	35.4	p<.001***
P03	8.9	p=.18 (n.s.)
P04	17.6	p<.001***
P05	-10.9	p=.18 (n.s.)
P06	16.1	p=.09 (n.s.)
P07	17.3	p<.001***
P08	17.7	p=.31
P09	39.1	p<.001***
P10	-0.4	p=.96 (n.s.)
P11	1.1	p=.59 (n.s.)
P12	70.1	p<.001***
P13	49.4	p<.001***
P14	17.4	p<.05*
P15	-16.5	p=.43 (n.s.)
P16	80.8	p<.001***
P17	84.3	p<.001***
P18	58.8	p<.001***
P19	8.6	p=.38 (n.s.)
P20	26.9	p<.05*
P21	114.8	p<.001***
P22	15.4	p<.01**
P23	44.9	p<.001***
P24	37.5	p<.001***
P25	13.4	p<.001***
P26	18.2	p=.31 (n.s.)
P27	41	p<.05*
P28	92.4	p<.001***
P29	-28.1	p=.36 (n.s.)
P30	36.9	p<.05*
P31	-4	p=.62 (n.s.)
P32	-11.6	p<.001***

Table 2.4. Differences of the individual differences between f0 maxima in S3 and S4 in the IP and word boundary (Δf0 IP# - Δf0 word#) with the results of the Simple Linear Regression models (LR results).

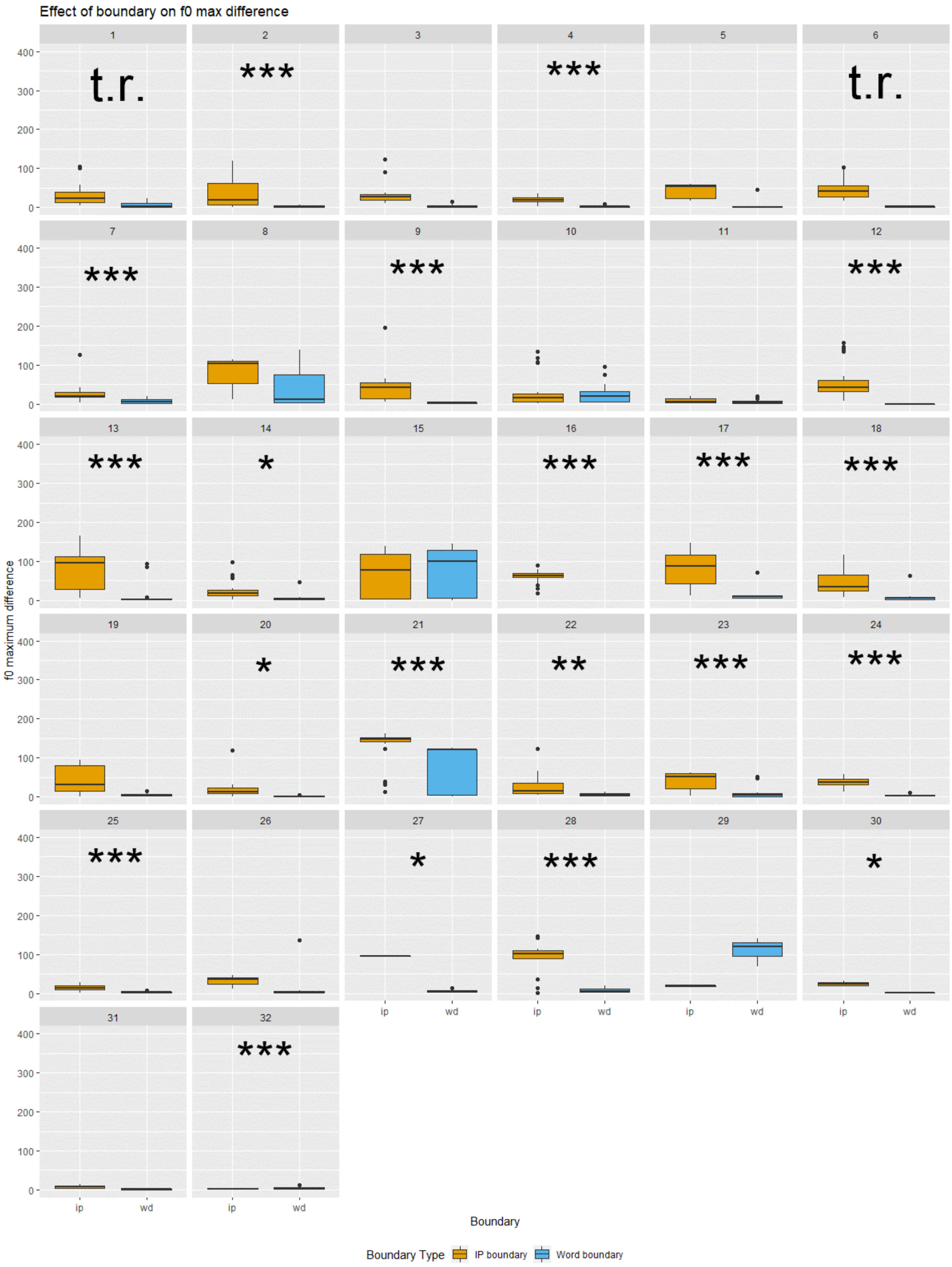


Figure 2.11. Effect of boundary type on f0 maximum difference for 32 individual participants. (***: p < .001; **: p <.01; *: p < .05; tr: p < .06)

The pitch reset values of five participants showed a significant effect of TG1 TYPE on Δf0.

Four out of these five participants produced significantly greater Δf0 when TG1 was 'maMIma' than when it was 'naNIna'. One participant (P15) showed the opposite pattern, producing significantly greater Δf0 when TG1 was 'naNIna'.

2.3.2.3. Pause duration

Pause duration was measured as the silent interval from the offset of the final syllable of the pre-boundary target word (TG1) to the onset of the first syllable of the post-boundary target word (TG2). All stimuli in the IP boundary condition were measured. Across the 32 individual participants, the average pause duration ranged from as short as 139ms to as long as 643ms, indicating substantial variation across individual speakers. Figure 2.12. shows average pause duration by participant.

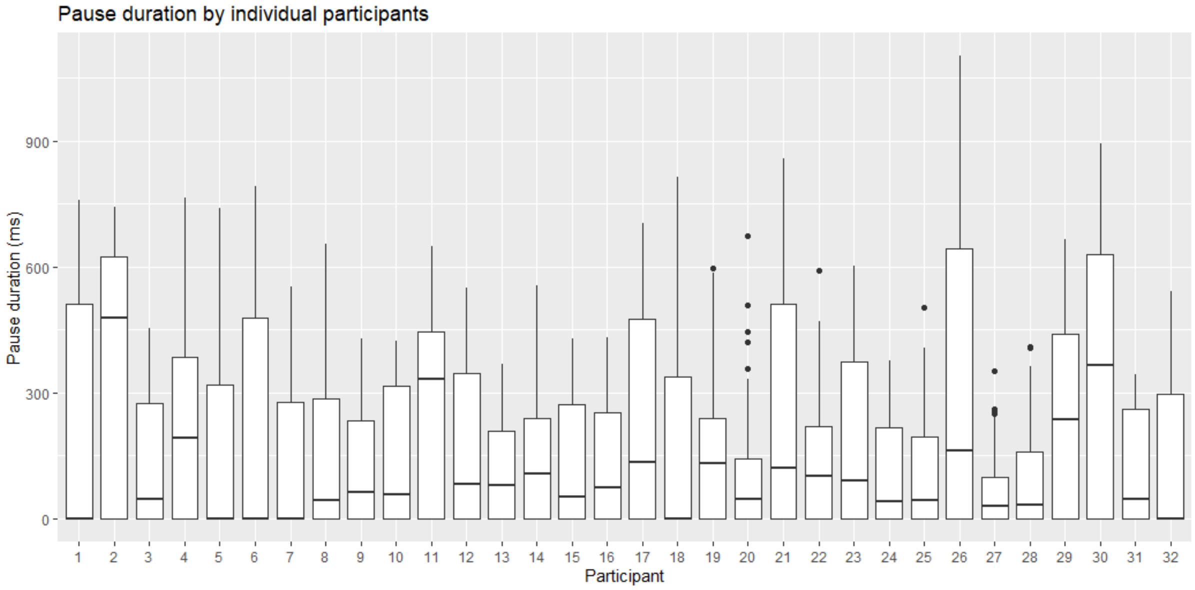


Figure 2.12. Average pause duration (ms) measured across IP boundary by participant.

2.3.3. Relationships between measurements

To investigate if there are systematic relationships between the different temporal properties and the temporal and tonal properties associated with the IP boundary, Correlation Tests were conducted in R. The correlations between S3 duration and pitch reset, S3 duration and pause duration, and pause duration and pitch reset were tested. The analysis was done for each TG1. Scatter plots with the regression line and the correlation coefficients for each of the three comparisons are shown in Figure 2.13.

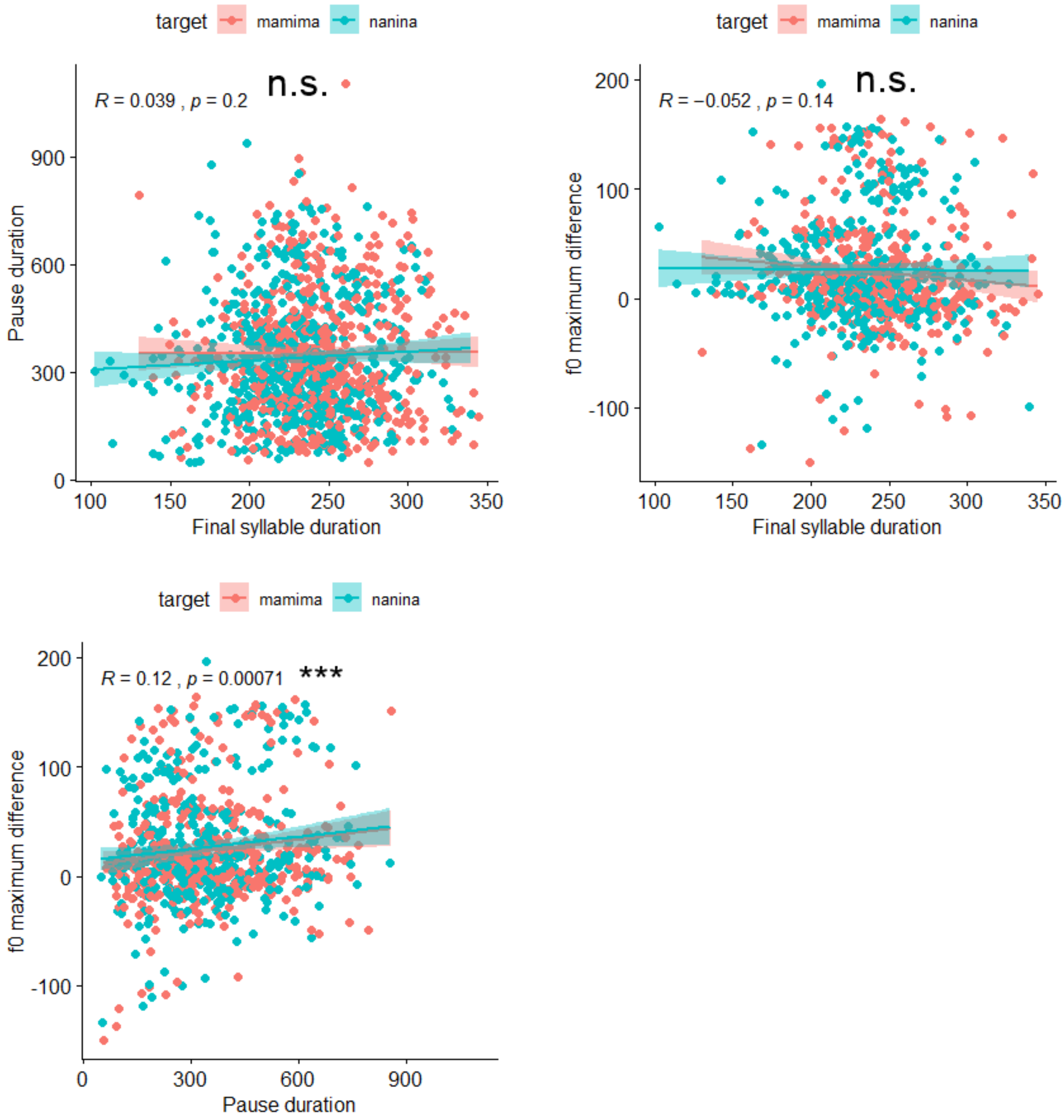


Figure 2.13. Scatter plots for each combination of the three acoustic properties of IP boundary. (***: p < .001; n.s.: p < .05)

When all productions of the stimuli in the IP boundary condition were pooled across participants, there is a significant positive correlation between pause duration and f0 maximum difference (p < .001; bottom panel in Figure 2.13), whereas the other two correlation coefficients were not significant. However, when the same correlation tests were conducted individually, different patterns emerged. The individual results are shown in Figures 2.14 through 2.16. The two correlations involving the f0 measure were not calculated for the two participants (P27 and P30) who did not have enough datapoints for the f0 measure in the IP boundary condition.

For S3 duration and pause duration (Figure 2.14), five out of 32 participants (P6, P10, P24, P29, P32) showed a significant negative correlation, meaning that the longer their pause durations were, the shorter their S3 durations. On the other hand, one participant (P1) showed a significant positive correlation, such that the longer their pause durations, the longer their S3 durations. For S3 duration and f0 max difference (Figure 2.15), four out of 30 participants (P13, P19, P21, P25) showed a significant negative correlation, and one participant (P14) trended in this direction (p = .051). The longer their S3 durations, the smaller their f0 max differences. The negative correlations are suggestive of these acoustic properties being in a trading relation with one another for a subset of participants.

For f0 max difference and pause duration (Figure 2.16), five out of 30 participants (P1, P4, P7, P13, P31) showed a significant positive correlation and one participant (P23) showed a trend of a positive correlation (p = .057). That is, the longer their pause durations, the greater their f0 maximum differences were across the IP boundary, thus showing an enhancing relation between the properties. For these speakers at least, stronger boundaries are marked by both temporal and tonal properties, with stronger boundaries resulting in larger pauses and larger pitch resets.

Figure 2.14. Correlation between pause duration (x-axis) and S3 duration (y-axis) for 32 individual participants. Six speakers showed a significant correlation. (***: p < .001; **: p <.01; *: p < .05; n.s.: p > .05)

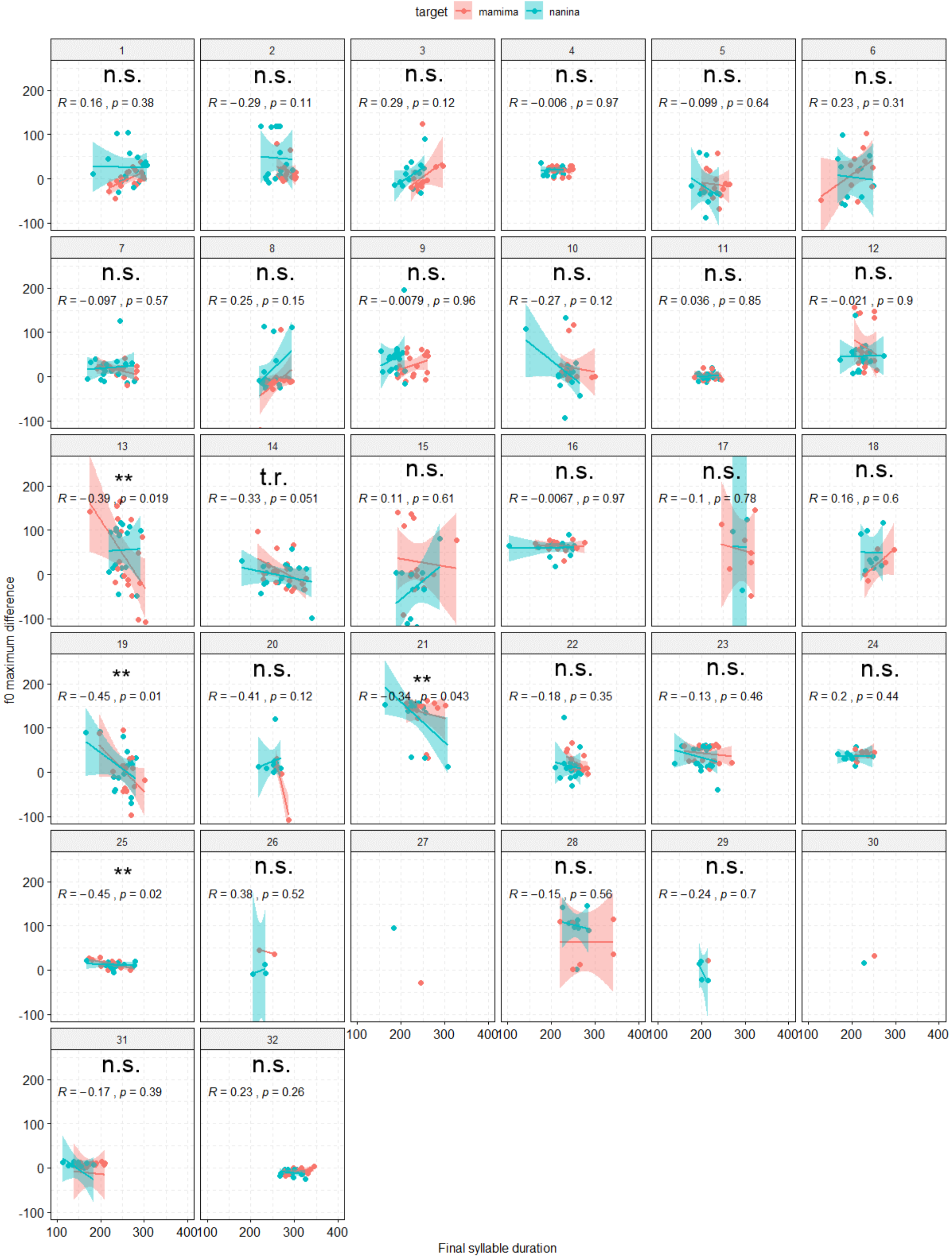


Figure 2.15. Correlation between S3 duration (x-axis) and f0 maximum difference (y-axis) for 32 individual participants. Four speakers showed a significant negative correlation, and one participant showed a trend in that direction. (***: p < .001; **: p <.01; *: p < .05; n.s.: p > .05)

Figure 2.16. Correlation between pause duration (x-axis) and f0 maximum difference (y-axis) for 32 individual participants. Five speakers showed a significant positive correlation, and one participant showed a trend in that direction. (***: p < .001; **: p <.01; *: p < .05; n.s.: p > .05)

2.4. Discussion

In this large-scale production study, durations of four syllables in the target word sequence, pause duration, and the f0 difference across prosodic boundary were tested to examine the effect of the IP boundary. Pause duration consistently occurred across the IP boundary for the 32 participants. Linear Mixed-effect models showed that, for data aggregated across participants, the durations of syllables adjacent to the prosodic boundary were subject to pre-boundary lengthening and post-boundary shortening, and the f0 difference was larger across the IP boundary than the word boundary. However, separate Regression analyses for individual participants revealed that there are multiple different patterns of temporal and tonal modification near prosodic boundaries.

This section discusses the acoustic results for the three properties of IP boundaries for individual speakers, and the implications of those results for theories of prosodic structure.

2.4.1. Boundary-related lengthening and shortening

Across 32 participants, pre-boundary lengthening was observed in the last two syllables (S2 and S3) of the trisyllabic TG1 ('maMIma' or 'naNIna'), and post-boundary shortening was observed in S4 of TG2. In addition, the current study finds six different patterns of boundary effects in the durations of the three syllables of TG1, shown in Figure 2.17. In Group 1, 15 out of 32 participants showed lengthening of the last two syllables (S2 and S3), consistent with the overall group pattern for these syllables. In Group 2, nine participants showed lengthening only of the last syllable (S3). In Group 3, three participants also showed lengthening confined to S3, but they also showed shortening of the first syllable (S1), without showing an effect of boundary on the stressed syllable (S2). The two participants in Group 4 lengthened all three syllables, whereas the two participants in Group 5 showed lengthening of S3 and shortening of S2. In Group 6, the one participant lengthened S2 and S3 but shortened S1.

The variation across speakers corroborates previous studies that reported individual differences for phrase-final lengthening (Fougeron & Keating, 1997; Byrd & Saltzman, 1998; Byrd et al., 2006; Mo & Cole, 2010), and adds to the body of research by demarcating the temporal modulation of the speaker-specific boundary effects.

Figure 2.17. Average durations of syllables of TG1 (left of the vertical line representing zero) and of first syllable of post-boundary word TG2 (right of the vertical line) produced in IP boundary condition by 32 participants. Pink, green, yellow: syllables with a lengthening, shortening, and no effect due to boundary, respectively.

The findings of robust boundary-related lengthening of the word-final syllable and leftward spread to the previous syllable(s) support previous studies that report lengthening of segment(s) preceding the phrase-final syllable (e.g., Berkovits, 1994, for Israeli Hebrew; Cambier-Langeveld, 2000, for Dutch; Fougeron & Keating, 1997; Shattuck-Hufnagel & Turk, 1998; Turk, 1999; Turk & Shattuck-Hufnagel, 2007 for English; Katsika, 2016 for Greek). The observed lengthening effect on the penultimate syllable which bears lexical stress in the trisyllabic target words ('maMIma' and 'naNIna') is consistent with previous studies showing a continuous and local scope of lengthening (Berkovits, 1994; Cambier-Langeveld, 1997; Byrd et al., 2006; Katsika, 2016)[1].

Individual variation is observed in the onset and the scope of boundary-related lengthening. In terms of the onset of the boundary effect in the pre-boundary target word (TG1), participants in Groups 1 and 6 produce earlier onset of lengthening beginning at S2 in TG1 than participants in Groups 2, 3, and 5, who lengthen only S3. The two participants in Group 4 produce an even earlier onset of lengthening, at S1.

The scope of the boundary effect also extends rightwards, affecting the duration of the first syllable of the post-boundary target word (S4).

Across all speakers, S4 duration was significantly shorter after the IP boundary than the word boundary. However, not all participants produced this shortening pattern. Six participants in each of Groups 1 and 2 produced post-IP boundary shortening. The two participants in Group 4 did not adjust S4 duration depending on the boundary type, while all participants in Groups 3, 5, 6 produced shorter S4 duration after the IP boundary than after the word boundary. Thus, a total of 18 participants produced S4 shortening, while the remaining 14 participants did not show an effect of boundary on S4 duration.

The boundary effects observed in the pre-boundary target word (TG1) are consistent with the predictions of the π-gesture model (Byrd & Saltzman 2003). First, a continuous lengthening in TG1 is found across the board, in line with predictions of the π-gesture model, where all consonant and vowel gestures that are active at the same time as the π-gesture are subject to the effect of the π-gesture and its effect of local slowing. In addition, this study found individual differences in the variable onset of lengthening (starting at S1, S2, or S3) and scope of phrase-final lengthening (one, two, or three syllables). With the exception of the two participants in Group 4, the patterns of pre-boundary lengthening observed in the 30

---

[1] The only exception to the continuous effect that I am aware of is a finding reported in Turk & Shattuck-Hufnagel (2007). They found a continuous effect for the same type of words (e.g., 'pacific', 'manassas') as analyzed in the present study, but did not find it for other trisyllabic words used in that study such as words with the main stress on the antepenultimate syllable (e.g., 'Madison', 'Cheddarfield).

participants is consistent with the account proposed in Katsika et al. (2014). Katsika et al. (2014) argued that the π-gesture and the gesture that models lexical stress (μ-gesture) are weakly coupled to each other phrase-finally. As a consequence of the interaction of the two gestures, the π-gesture shifts to the left towards the stressed syllable, resulting in the earlier onset of phrase-final lengthening in words with non-final lexical stress. This account suggests that boundary effects can spread towards the stressed syllable. The differences between speakers observed in the current study might be attributed to variation in the coupling of the π-gesture with other gestures (for example, the variable strength of the coupling between π-gesture and the μ-gesture resulting in the variable degree in which the π-gesture shifts towards the stressed syllable), and/or the variable scope of the π-gesture itself. It should be noted that a more fine-grained analysis of the data, such as an analysis of the segment durations depending on the boundary, as well as computational modeling of the data are required to evaluate to what extent the individual differences observed in the results are due to systematic variation between speakers.

Turning to the shortening effects, post-boundary shortening was found for S4 (across speakers and for 18 speakers in the individual analyses). The pre-boundary target word showed shortening of S1 or S2 durations for a subset of participants as well. The four participants in groups 3 and 6 shortened S1 duration in the IP boundary condition, and the two participants in group 5 shortened S2 duration in the IP boundary condition. Not many previous studies have identified boundary-related shortening (though see Byrd et al., 2006; Krivokapić, 2007; Katsika et al., 2014). Similar to the present study, Katsika et al. (2014, see also Katsika, 2016) found both pre- and post-boundary shortening effects and interpreted them to be a consequence of the interaction between lexical stress and boundary-related lengthening (rather than a separate effect of the boundary). In Katsika's model, the interaction between the μ-gesture in the pre-boundary word and the π-gesture results in not only shifting of the π-gesture (as explained above) but also shifting of the μ-gesture towards the boundary. That is, the π-gesture shifts towards the lexical stress in the pre-boundary word, which leads to post-boundary shortening, and the lengthening effect of the lexical stress shifts towards the boundary, which leads to shortening at the onset of stress-induced lengthening. This account can capture most of the patterns observed in the present study, with the differences between individual participants being possibly due to the result of the variable scope and/or coupling strength of the π-gesture and the μ-gesture.

2.4.2. *Pitch reset*

A positive pitch reset in the IP boundary condition indicates that there was a significant increase in the f0 in the post-boundary syllable compared to the f0 in the pre-boundary syllable. That a majority of participants (20 out of 32 participants) used f0 to mark the IP boundary suggests that tonal patterns associated with the right edge of IP (the pitch reset measure in this study indirectly captures the end of the L boundary tone associated with the stressed deaccented syllable of TG1) are used in conjunction with the temporal properties (Ladd, 1988; de Pijper & Sanderman, 1994). However, 11 participants produced an f0 difference across word boundaries that was not significantly different from the f0 difference produced across IP boundaries. The results suggest that these participants did not employ pitch reset to distinguish IP boundaries from word boundaries. There was one participant who produced a small positive reset at word boundaries but a negative difference at IP boundaries, which is an unexpected pattern.

Those speakers who used f0 to signal IP boundaries produced f0 differences across IP boundaries to varying degrees. For example, some participants – e.g., participants 4, 6, 14, 24, 25, 30, 31, 32 – produced an average f0 difference across IP boundaries smaller than 50Hz, while other participants – e.g., participants 1, 2, 20, 21, 22, 28 – produced average f0 difference across IP boundaries that exceeded 100Hz. Individual variation in the use and degree of f0 in marking IP boundaries is in line with a few previous studies that showed substantial individual variation in the f0 contours associated with a major phrase boundary (Swerts, 1997; Zhang, 2012; Petrone et al., 2017). In Petrone et al. (2017), the majority of speakers produced a H edge tone to signal the end of IP in German, but the 12 individuals differed in the type of f0 contours, such as rise, plateau, fall or others. Petrone et al. (2017) concluded that f0 is the most variable property of IP boundaries due to the robust individual variation. In Zhang (2012), the 10 English speakers showed substantial variation in the mean f0 slope, with two speakers showing only a very small f0 rise across the IP boundary, suggesting that they did not employ the pitch cue to signal the IP boundary. The results of the current study also indicate pitch reset to be the most inconsistent property in that not all participants produced a positive pitch reset across IP boundaries, whereas all participants produced final lengthening of the word-final syllable and pauses at IP boundaries.

2.4.3. *Pause duration*

Inter-speaker variation in pause duration has been relatively well documented in previous studies (Swerts and Geluykens, 1994; Fant et al., 2003; Krivokapić, 2007). The results of the current study are consistent with these studies, showing substantial variation in the duration of silent interval produced at the IP boundaries.

It has been argued that, as with other properties of IP boundaries, the high variability in pause duration may be correlated with the strength of the prosodic boundary (Swerts, 1997, Krivokapić, 2007, Horne et al., 2015). Boundary strength is unlikely to be a major cause of the large variability found in the current study given that the stimuli did not differ across participants and participants were given the same instructions, though fine aspects of interpretation might have differed between speakers. Therefore, the variation in pause durations presented here may be predominantly the result of individual differences in boundary production. Petrone et al. (2017) noted that durations of silent pauses for IP boundaries were perceived in a categorical manner, despite the variability in production. It is worth investigating further how listeners deal with the inter-speaker variability in the production of IP boundaries and whether and how perception of pause duration differs from perception of other properties of the IP boundary, which are questions that are addressed in the next chapter of this dissertation.

2.4.4. Weighting of the acoustic properties of IP boundary

Figure 2.18 summarizes the results for the three types of IP boundary markers by giving each speakers' averaged produced syllable durations (left), pause duration (middle), and pitch reset values (right). As discussed, not all speakers use all three acoustic properties to distinguish IP from word boundaries, and speakers differ in the degree to which they use the acoustic properties. Thus, 11 participants did not significantly differ in pitch reset across IP vs. word boundaries, and one participant produced a difference but in a direction opposite the expected one. Among those 20 participants who did use pitch reset, some produced a larger f0 difference across IP boundaries than others. For the other two examined properties, participants used them to varying degrees.

The results do not indicate a predominant pattern among the 32 participants in how the various acoustic properties relate to each other. That is, it is not generally the case that speakers who use a specific property to a larger (or lesser) degree use one or more of the other properties to a larger (or lesser) degree as well.

Nonetheless, despite the absence of these general patterns, some participants manifested trading relationships for some of the markers, as shown by the results of the Correlation Tests for S3 duration and pause duration (five participants) and for S3 duration and f0 difference (five participants who did not overlap with those first five participants).

This means that the more these participants produced one property of IP boundary, the weaker their production of the other property. The trading relation between S3 duration and pause duration is in line with previous studies (Ferreira, 1993; Byrd et al., 2006).

On the other hand, for some participants, the results for pause duration showed an enhancement relationship relative to other properties. Six participants showed this relationship for pause and pitch reset: the longer their pause durations, the greater the f0 difference across IP boundaries. One participant showed this relationship between pause duration and S3 duration, such that the longer their pause durations were, the shorter the S3 duration.

Overall, though, it was not the case that either a trading or enhancement relationship was observed across all participants, meaning that there was no single overarching way that participants use one marker of IP boundary in relation to how they use another. The precise relationship between the various IP boundary markers needs to be further investigated.

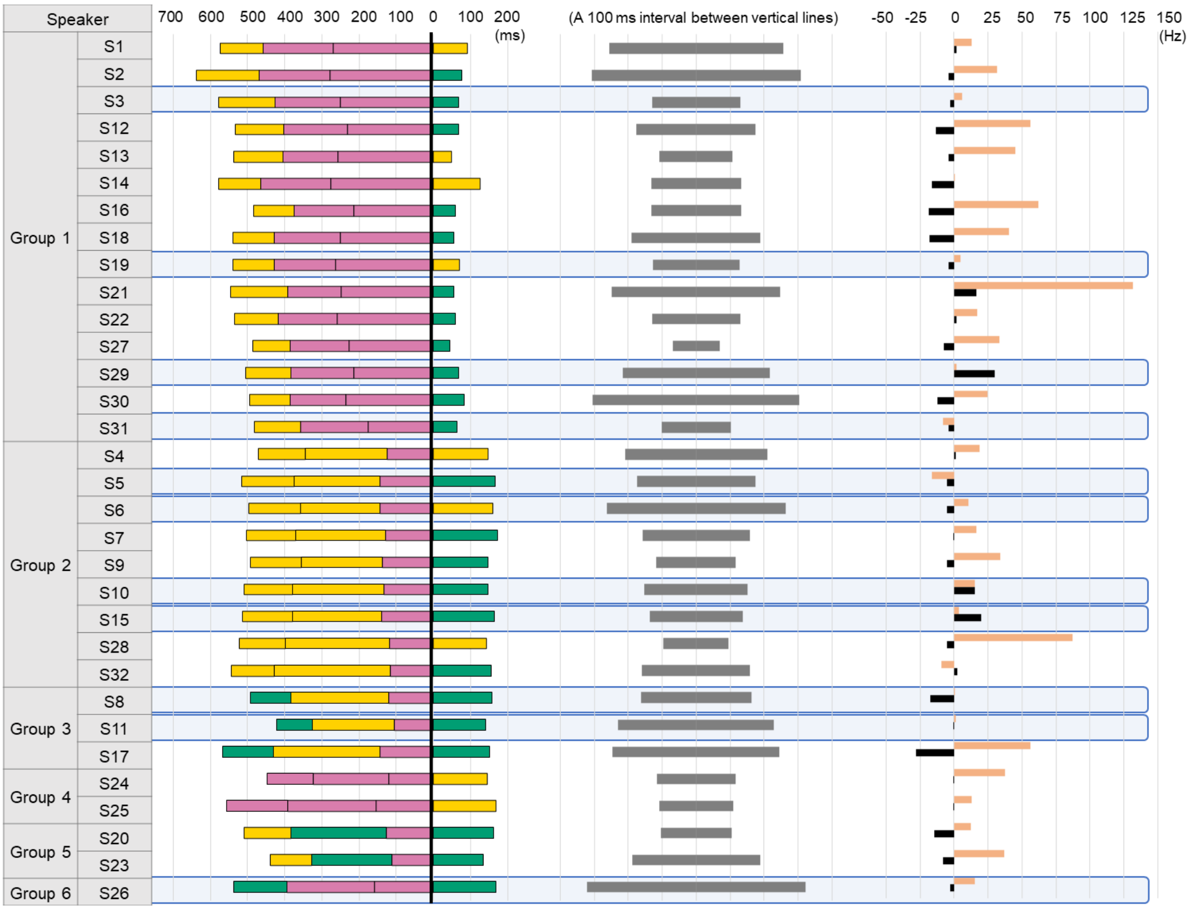


Figure 2.18. Results of analyses of acoustic properties associated with IP boundaries for 32 participants. Left: the average durations of each syllable of TG1 and of the first syllable of TG2 in IP condition. Middle: average pause durations. Right: average f0 difference in word (black) and IP (orange) boundary conditions. Participants marked with the blue box did not use pitch reset to distinguish IP and word boundaries.

2.5. Conclusion

The results of the production study that analyzed 32 participants' production of IP and word boundaries showed that all three acoustic properties are reliably used to distinguish IP from word boundaries. Individual-speaker analyses revealed that there is substantial variation in both the combination of the IP boundary markers and the degree to which they are employed. Although the study found continuous effects of IP boundaries on boundary-related lengthening, the variation in the onset and scope of the lengthening among participants and the differences in how they use pitch reset suggests that individuals differ in how they encode prosodic structure.

Chapter 3 presents the perception experiment that investigated whether the individual variation observed in the production study is mirrored in the same participants' perception.

CHAPTER 3

Perception of Prosodic Boundaries

In this chapter, American English-speaking listeners' real-time use of the acoustic properties that differentiate IP and word boundaries is investigated using a visual world paradigm. Particular attention is paid to the perceptual weighting of the multiple acoustic correlates identified in Chapter 2. Section 3.1 introduces the details of the methodology used in the experiment including the creation of the auditory and visual stimuli, participants, procedures, and statistical design. Section 3.2 presents the hypotheses and predictions of the perception experiment. Section 3.3 reports the results of the statistical analyses conducted based on the group data. Finally, Section 3.4 discusses the results in regard to the three major cues for IP boundary.

3.1. Methods

3.1.1. Auditory stimuli

The experimental sentences from the production experiment were used in the perception experiment. Auditory stimuli were created by manipulating the acoustic properties of the target word sequence (TG1 # TG2, where # denotes either IP or word boundary) in the target sentences as produced by the model speaker.

The model speaker was a female native speaker of American English from the Midwest. The procedures for recording the model speaker were the same as those used for the production experiment described in Chapter 2. The model speaker was seated in front of a microphone and a laptop computer inside a sound-attenuated booth and given the same written instructions as the participants in the production experiment. The recording consisted of four repetitions of the eight pairs of target sentences. The model speaker was given the same stimuli and instructions as the participants for the production experiment, and therefore silently read the context sentences before saying the target sentences out loud.

Table 3.1 gives the model sentences that served as the basis of the acoustic manipulations. They differ from the actual recording in that the first word of each target sentence ('no') was removed to reduce the time per each trial.

TG1	TG2	Boundary	Auditory stimuli
maMIma	Melinda	IP	the **police** called maMIma. # Melinda and **Danny** said no one got hurt.
		Word	the **police** called maMIma # Melinda. And **Danny** said no one got hurt.
	Belinda	IP	the **queen** called maMIma. # Belinda and **Daisy** thought that was rude.
		Word	the **queen** called maMIma # Belinda. And **Daisy** thought that was rude.
naNIna	Navarro	IP	the **painter** called naNIna. # Navarro and **Damon** bought the painting.
		Word	the **painter** called naNIna # Navarro. And **Damon** bought the painting.
	Delilah	IP	the **farmer** called naNIna. # Delilah and **David** asked about the apples.
		Word	the **farmer** called naNIna # Delilah. And **David** asked about the apples.

Table 3.1. Target sentences used in the perception experiment.

The auditory stimuli were created by manipulating syllable durations, pause duration, and f0. Prior to these manipulations, the following measures were taken for all sentences produced by the model speaker: the durations of all three syllables of TG1 and the first syllable of TG2, the silent interval (pause) between TG1 and TG2 in the IP boundary condition, and the f0 minimum and maximum in the last syllable of TG1 (S3) and the first syllable of TG2 (S4). The averages of these values were calculated and used to determine the values of the manipulated auditory stimuli.

3.1.1.1. Manipulation procedures

This section describes the manipulation procedures for each acoustic property to create the auditory stimuli containing a specific set of IP boundary markers. For each of the four IP vs. word comparisons (see Table 3.1), the same base carrier utterance (produced in the word boundary condition) was used across different manipulations. Consequently, there is no difference in the auditory stimuli before and after the target word sequence within the same sentence type.

a) *Pause*

There were two manipulation conditions for pause: present and absent. In the pause-present condition, the model speaker's average IP pause duration (240 ms) was inserted between TG1 and TG2. In the pause-absent condition, the same pause duration was inserted after TG2, while TG1 and TG2 were concatenated without a silent interval.

b) Pitch reset

There were three pitch reset conditions, as schematized in Figure 3.2. Figure 3.3 gives an example of the pitch reset manipulation with the target word sequence “maMIma # Belinda”. The “No pitch reset” manipulation is based on an interpolation between the two f0 values at S3 onset and S4 offset after removing all f0 values between these values. This achieved a gradual decrease in f0 across the boundary. The difference between the two f0 targets was less than 10 Hz, which is unlikely to be perceived as a f0 fall for speakers of American English (Turner, Bradlow & Cole, 2019).

The “Small pitch reset” manipulation is based on the model speaker’s average difference between the f0 minima in the two syllables across IP boundary (S3 and S4), which was 21 Hz. (The average difference between the f0 maxima across IP boundary was 19 Hz. In the interest of a robust manipulation, the larger of the two values was chosen.) The “Large pitch reset” manipulation was 32Hz, which is the reference (i.e., “Small pitch reset”) size multiplied by 1.5. All f0 manipulations were applied after the syllable durations of TG1 were manipulated and concatenated.

The first step of applying the two different sizes of pitch reset to the different versions of TG1s was to set the f0 target at the offset of TG1 at 161Hz, which was held constant across the different manipulation conditions. This f0 target was then interpolated with the f0 target at the onset of TG1, and this interpolation created a f0 trajectory that was comparable to a LL%. This f0 trajectory for TG1 ensured that participants were not exposed to any f0-related variation leading up to the target boundary. Then the f0 target at the onset of TG2 was set to 182Hz (i.e., 21Hz higher than 161Hz) for the manipulation types using the reference size of pitch reset (“Small pitch reset”), and to 193Hz (i.e., 32Hz higher than 161Hz) for the manipulation type using the larger size of pitch reset (“Large pitch reset”). For the control conditions, the f0 target at the onset of TG2 was set to 161Hz.

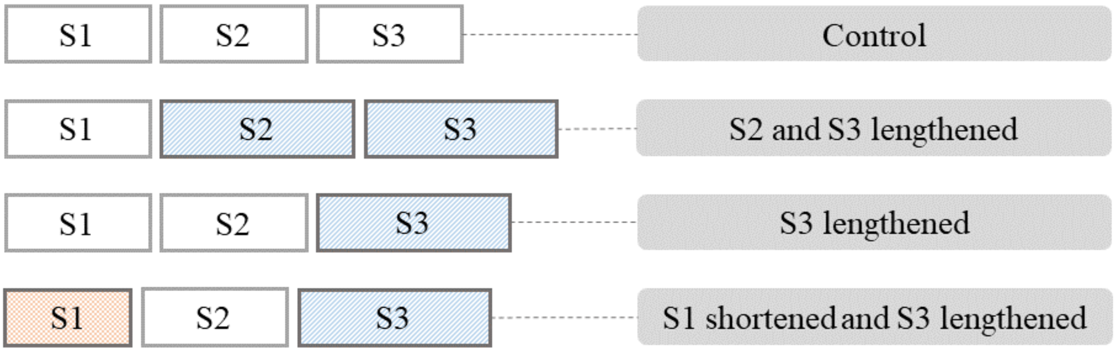


Figure 3.1. Schematic representation of the manipulation conditions for phrase-final lengthening.

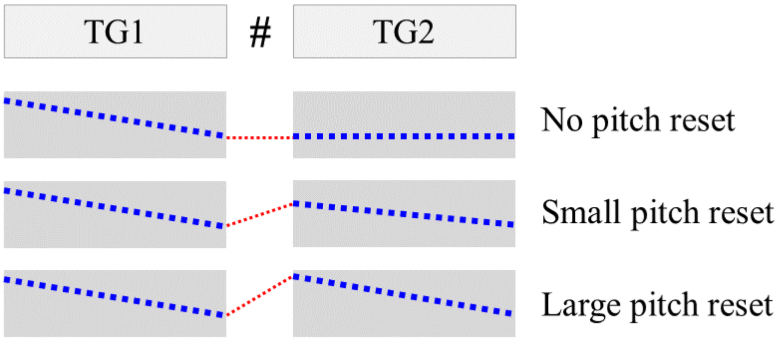


Figure 3.2. Schematic representation of the manipulation conditions for pitch reset.

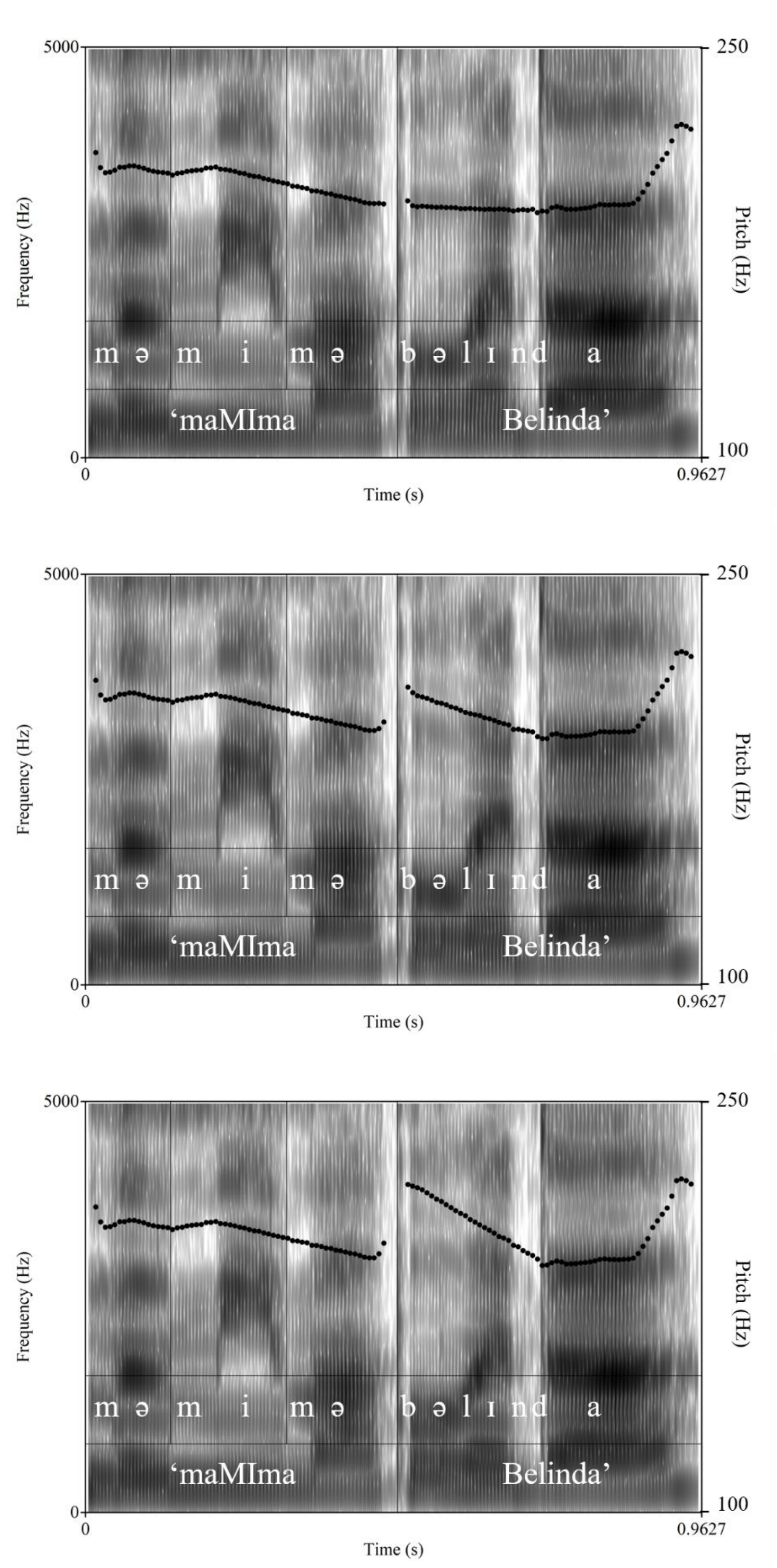


Figure 3.3. Spectrograms of the target word sequences "maMIma Belinda" for the different pitch reset conditions (top: no reset, middle: small reset, bottom: large reset) used in the stimuli.

c) Final lengthening

As shown in Table 3.2, there were four syllable duration conditions, with the manipulated durations being determined by the model speaker's TG1 ('maMIma', 'naNIna') syllable durations in the word and IP boundary conditions averaged across repetitions. The syllable durations used in the control condition are the average syllable durations produced in the word boundary condition. Shortening and lengthening of syllable length was done by removing (for shortening) or adding (for lengthening) pulses at the end of the syllable.

Lengthening manipulation	TG1 = 'maMIma'			TG1 = 'naNIna'		
	S1	S2	S3	S1	S2	S3
Control	137	175	177	131	172	160
S2 and S3 lengthened (Early onset lengthening)	137	211	309	131	207	302
S3 lengthened (Late onset lengthening)	137	175	309	131	172	302
S1 shortened and S3 lengthened (Late onset lengthening + shortening in S1)	110	175	309	105	172	302

Table 3.2. The manipulated syllable durations (ms) for each TG1.

The four manipulation conditions of final lengthening are schematized in Figure 3.1. The control condition reflects the average syllable durations of TG1 produced in the word boundary condition by the model speaker. The three remaining conditions reflect the three most common patterns of phrase-final lengthening observed in the production study. The durations of the lengthened and shortened syllables are derived from the speaker's averaged productions of TG1 in the IP boundary condition. The three syllables (S1, S2, and S3) of TG1 were spliced from three different renditions of TG1 that matched the average syllable duration for the control condition.

3.1.1.2. Summary of manipulation conditions

The model speaker's original, unmanipulated utterances produced in the IP boundary condition manifested all three acoustic cues to IP boundary: phrase-final lengthening, pause, and pitch reset. The goal of the artificial manipulation of these acoustic markers of IP boundary was to control the combination and type of the markers available in the auditory stimuli. The two manipulation conditions of pause, four manipulation conditions of syllable durations of TG1 including a control condition, and three manipulation conditions of pitch reset resulted in 12 different combinations of manipulation conditions, which are represented in Figure 3.4.

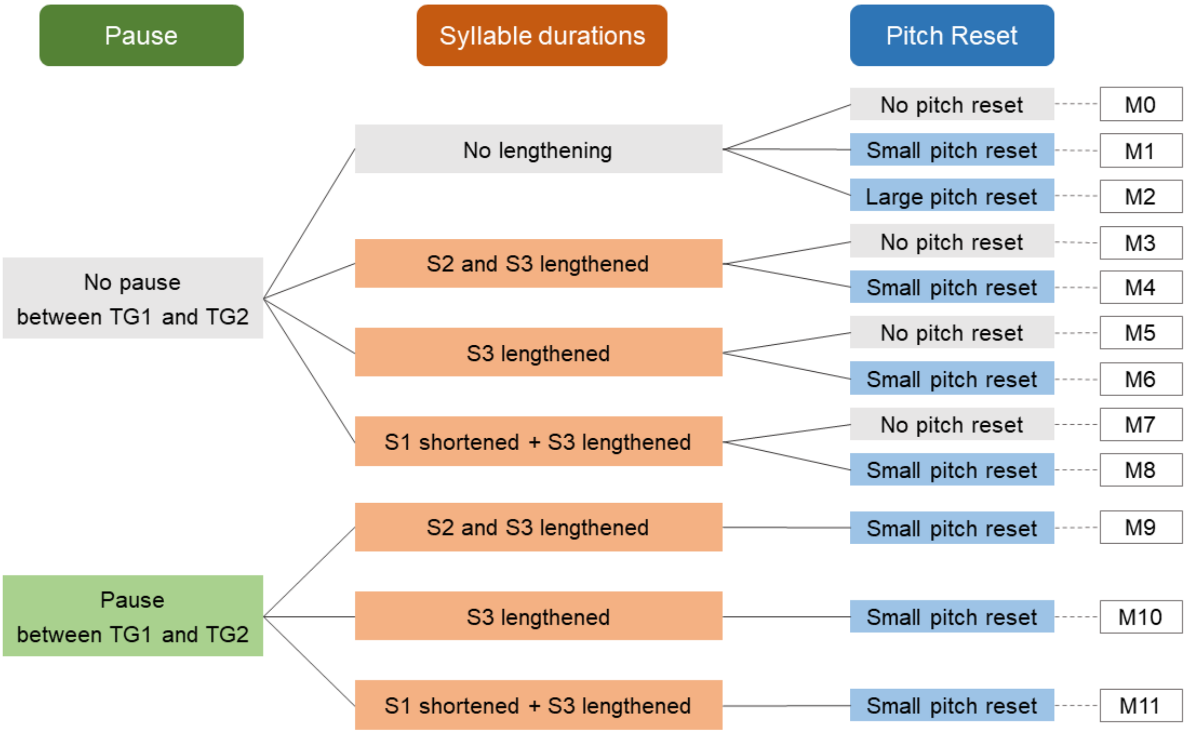


Figure 3.4. Summary of all manipulation conditions. M0~M11 refer to the combinations of the three IP boundary markers.

Pause is inserted after TG2 in nine out of the 12 manipulation conditions. Three out of those nine conditions do not involve boundary-related lengthening or shortening, such that the syllable durations of TG1 reflect the model speaker's averages produced in the word boundary condition. These three conditions are distinguished in the manipulation of pitch reset; one does not involve pitch reset, another involves a small pitch reset between TG1 and TG2, and the other involves a large pitch reset. The condition without any of the three IP boundary markers serves as the control condition (M0), while the other two conditions (M1, M2) test the effect of the size of pitch reset. Varying the size of pitch reset is motivated by the gradient sizes of pitch reset found in the participants' acoustic data collected in the production experiment.

The remaining six out of the nine conditions are characterized by three different patterns of phrase-final lengthening and the presence/absence of pitch reset. In the absence of pause and pitch reset between TG1 and TG2, the three conditions differing only in the pattern of lengthening (M3, M5, M7) allow us to compare the timing of perception of an IP boundary based solely on the different time points at which a temporal cue – syllable shortening or lengthening – is heard. The point of investigating the other three conditions without pause but with pitch reset (M4, M6, M8) is to determine how listeners respond to the combined effect of final lengthening and pitch reset. Previous research on cue-weighting of prosodic boundaries has examined whether the effect of these cues is cumulative (Streeter, 1978; Seidl, 2007; Yang et al., 2014) – i.e., whether listeners would detect an IP boundary more accurately when given a combination of two cues than when given a single cue. While Streeter (1978) and Seidl (2007) found evidence for cumulative effects of acoustic cues to perception of boundary in English and German respectively, Yang et al. (2014) did not find such effects in Mandarin.

Pause is inserted after TG1 in the remaining three manipulation conditions (M9, M10, M11). All three conditions have a small pitch reset across the target boundary, and they vary depending on the pattern of phrase-final lengthening. Because listeners will have access to final lengthening first, then pause, and then pitch reset, the time course of listeners' fixations on the target visual stimuli may vary across these manipulations.

There are two main reasons that the conditions are not fully balanced – i.e., not all combinations of the cues are included in the design. First, the manipulated stimuli involving no final lengthening in the phrase-final TG1 followed by a silent interval did not sound natural (presumably because such utterances do not typically occur in speech production). The other reason for the imbalance is to reduce the length of the eye-tracking experiment by including only those manipulation conditions for which clear hypotheses can be established.

3.1.2. Visual stimuli

Eight pairs of simple line drawings were created by a professional artist to illustrate the critical portion of the situation described in each pair of target sentences. In the drawings, the person facing forward was always the person doing the "calling" (see Table 3.1): a policeman, a farmer, a queen, or a painter. The person with their back towards the viewer – referred to as either 'Mamima' or 'Nanina' – appeared the same in all pictures. The two pictures in each pair were differentiated by the presence/absence of a third person inside a speech bubble, which signaled the meaning difference induced by the type of prosodic boundaries. Figure 3.5 represents an example pair of images depicting the target sentences with the IP boundary condition on the left and the word boundary condition on the right. The left image without the speech bubble depicts a situation in which the painter called Nanina, while the right image with the speech bubble depicts the painter mistakenly calling Nanina using a different name, Navarro. In all pictures showing three people, the third person appeared the same. The image was gender neutral given that it variably referred to someone with either a typically female name (Melinda, Belinda, and Delilah) or a typically male name (Navarro). All pictures used in the perception experiment are given in Appendix F.

(1) IP boundary condition

(2) Word boundary condition

(1)

A: The agent called Nanina.   Navarro and Parker bought the painting.

B: No, the painter called Nanina. # Navarro and Damon bought the painting.

(2)

A: The agent called Nanina Navarro. And Parker bought the painting.

B: No, the painter called Nanina # Navarro. And Damon bought the painting.

Figure 3.5. Examples of visual stimuli with corresponding sentences. The pound sign (#) represents the location of the target boundary, IP in (1), word in (2). Participants heard the underlined portion of the sentences.

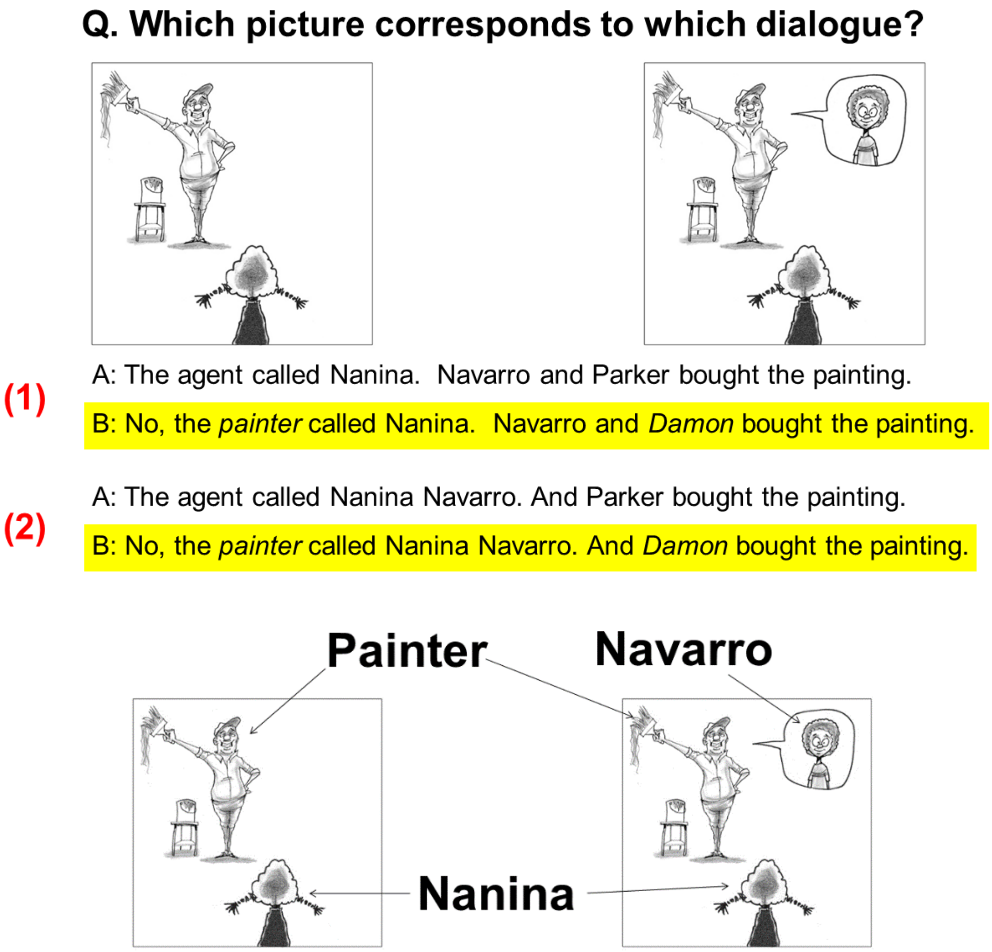
3.1.3. Participants

All 32 participants in the production experiment were invited to return to participate in the perception experiment. Twenty of them returned, with the separation between their participation in two experiments being six months or longer. They were tested individually and received financial compensation for participating in the roughly 75-minute perception session.

3.1.4. Procedures

After reading the instructions for the experiment and giving written consent to participate, participants were first seated in front of a computer monitor for a familiarization session. They were presented with a slideshow of the visual stimuli accompanied by the written version of the auditory stimuli. An example of these slides is shown in Figure 3.6. The location of the two images on the screen was randomized in the familiarization session as well as in the main test. Participants were told that they would not hear both sentences during the main experiment, but instead would hear only the highlighted portion (excluding the initial 'no'). They were told that the italicized words represent the words that are emphasized. They were reminded of the fact that the novel names 'maMIma' and 'naNIna' are interpreted as personal names.

The familiarization session ensured that participants establish associations between the picture and boundary type prior to testing. To check their understanding, after being given sufficient time to look at, for example, the top pair of pictures and dialogue in Figure 3.6, participants were asked to match the pictures with the sentences, and then match the people in each picture with the names in each pair of sentences. After answering these questions, they saw the next slide – for example, the bottom pair of pictures in Figure 3.6 – which showed the same pictures with the characters' names labeled.



Q. Which picture corresponds to which dialogue?

(1) A: The agent called Nanina. Navarro and Parker bought the painting.
B: No, the *painter* called Nanina. Navarro and *Damon* bought the painting.

(2) A: The agent called Nanina Navarro. And Parker bought the painting.
B: No, the *painter* called Nanina Navarro. And *Damon* bought the painting.

Figure 3.6. Sample familiarization slides. Participants saw the top slide first and then saw the bottom slide.

After the familiarization session, participants were seated in front of a computer monitor and a desk-mounted camera, with a distance of about 550-650 mm from the camera and about 800 mm from the monitor. Their eye movements were captured with a remote monocular eye-tracker (EyeLink 1000 Plus, SR Research), using a 25 mm lens and sampling at 500 Hz. At the beginning of the eye-tracking session, the experimenter performed a calibration procedure with the participant's dominant eye. It was repeated until criterion was reached. The auditory and visual stimuli were presented to the participants using Experiment Builder (SR Research) and professional headphones (AKG271MK2).

After calibration and before the main testing, participants were given 10 practice trials that were randomly chosen from the main test, to ensure they fully understand the task. In each trial, participants saw a pair of pictures on the computer monitor while hearing the recorded prompt ("Look at the pictures"). They had 2000 ms to look at the pictures after the prompt ended. After that, a third picture appeared in the center of the screen, as participants heard another prompt ("Look at the center"). After 800 ms, the auditory stimulus was played. The center image remained throughout the rest of the trial. Participants were instructed to look at the image that represents the dialogue that they think they hear. (There was no other input method such as mouse clicking or button pressing.) An example of the visual stimuli with the third image in the center is given in Figure 3.7.

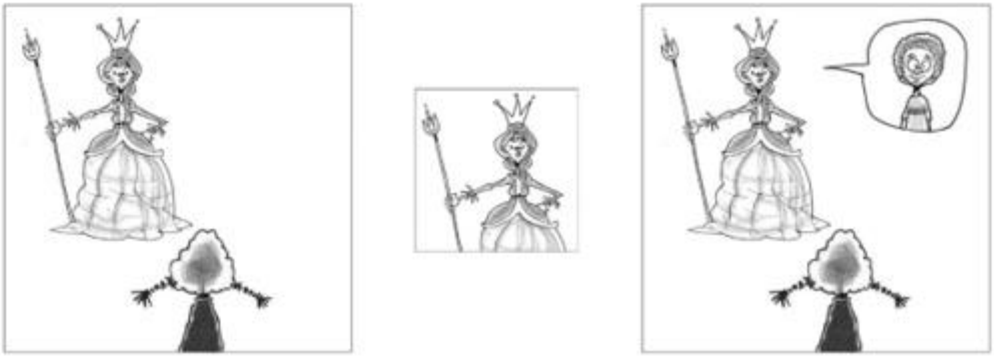


Figure 3.7. Sample trial screen.

The experiment included a break between each of four blocks. Each block consisted of one randomized repetition of the 48 auditory stimuli (two TG1 x two TG2 x 12 manipulation conditions), resulting in a total of 192 test trials for each participant. Each trial lasted no more than 10 seconds, and each block took less than nine minutes.

3.1.5. Analyses

The three images on the screen were defined as the Interest Areas (IAs) in which eye movements were monitored during each trial. In the binning analysis, proportion fixations to each IA over time was calculated for each 20ms temporal bin during each trial. For each trial, proportion fixations to all three pictures were calculated from the onset to the offset of the auditory stimulus. The 48 auditory stimuli had different durations due to the eight different sentence types (2 BOUNDARY x 2 TG1 x 2 TG2) and the different length manipulations. Due to the length manipulations, the time points corresponding to the onset of each syllable of TG1 also differed across conditions, as shown in Table 3.3 for the four different target sequences.

(1) maMIma # Belinda

Event	Time points (ms)			
	Control	Early onset lengthening	Late onset lengthening	Late onset lengthening + S1 shortening
Onset of TG1	1157	1157	1157	1157
Onset of S2 of TG1	1294	1294	1294	1267
Onset of S3 of TG1	1470	1505	1470	1442
Offset of TG1	1647	1814	1778	1751

(2) maMIma # Melinda

Event	Time points (ms)			
	Control	Early onset lengthening	Late onset lengthening	Late onset lengthening + S1 shortening
Onset of TG1	1132	1132	1132	1132
Onset of S2 of TG1	1270	1270	1270	1242
Onset of S3 of TG1	1445	1480	1445	1417
Offset of TG1	1622	1789	1754	1726

(3) naNIna # Delilah

Event	Time points (ms)			
	Control	Early onset lengthening	Late onset lengthening	Late onset lengthening + S1 shortening
Onset of TG1	944	944	944	944
Onset of S2 of TG1	1075	1075	1075	1049
Onset of S3 of TG1	1248	1282	1248	1222
Offset of TG1	1408	1584	1549	1523

(4) naNIna # Delilah

Event	Time points (ms)			
	Control	Early onset lengthening	Late onset lengthening	Late onset lengthening + S1 shortening
Onset of TG1	990	990	990	990
Onset of S2 of TG1	1121	1121	1121	1095
Onset of S3 of TG1	1293	1328	1293	1267
Offset of TG1	1453	1630	1595	1569

Table 3.3. Time points (in ms from 0 = onset of auditory stimulus) of TG1 syllables for target word sequences for different lengthening manipulation conditions and the control condition.

Participants’ eye movements were recorded and processed using EyeLink Data Viewer, a data analysis program that generated reports on proportion fixation on the three Interest Areas (IAs) using 20ms bins. For example, the proportion fixation on IA1 for a given time bin was calculated as the proportion of recorded fixation counts on IA1 divided by recorded fixations on all three IAs. Because IA1 and IA2 corresponded to the left and right images, respectively, while IA3 corresponded to the center image, which participants looked away from as soon as the auditory stimulus started playing in the trial, the proportions fixation on IA3 was only used to verify that participants were looking at the center image immediately before looking to the left or right image, and was not submitted to statistical analyses.

Data from one participant (P19) were excluded from the analyses because of difficulties in systematically tracking that participants’ eye gaze over the course of the experiment.

The eye movement data from 19 participants were analyzed using Generalized Additive Mixed Models (GAMM; Wood, 2006) using the mgcv package (version 1.8.28; Wood, 2011) in R (version 3.5.1; R core team, 2018; www.r-project.org). GAMM is a type of Generalized Linear Mixed-effects Model (GLMM) that uses non-linear smoothing functions to capture the non-linear relationship between two or more predictor variables, and therefore useful for fitting noisy data such as eye movement data.

Twelve manipulation types correspond to 12 different combinations of the three primary acoustic cues for the IP boundary (see Section 3.1.1.2). These 12 types were not tested in a single statistical model; rather, a subset of the perception data was modeled to examine the effect of each IP boundary cue in question. For example, responses to the four manipulation types involving the three patterns of phrase-final lengthening plus the control condition, all other cues being equal, will allow us to narrow down the question to see whether there are systematic perceptual differences in regard to the final lengthening patterns, when all other manipulation are held constant.

The significance testing of the effect of the manipulation types (i.e., the fixed effect of Pitch Reset or Lengthening Pattern for each model) was conducted via a nested model: a nested model including the parametric term (e.g., manipulation type) is set up and compared with a base model without the parametric term (and the difference smooth) using Akaike Information Criterion (AIC) scores, which take into account both goodness-of-fit and model complexity/simplicity. Model comparisons were conducted using the compareML function in the itsadug package (version 2.3; van Rij, et al., 2020) in R. The plots are created using the itsadug package and the ggplot2 package (version 3.2.1; Wickham, 2016)

The predictor variables relevant to the research questions were tested for significance. The response variables were the proportion fixations on the images representing either the IP boundary or the word boundary, that were labeled as ‘Proportion fixations to IP#’ and ‘Proportion fixations to word#’, respectively, in the analyses. Finally, as in the analysis of the production data, no corrections were made for multiple comparisons, with the idea that, given that both Type I and Type II error are errors, I rather show all results, with the understanding that this might increase the risk of Type I error.

3.2. General hypotheses

The experiment investigates whether there are significant differences in participants’ responses to the 12 manipulation types that represent 12 different combinations of the IP boundary cues. It is hypothesized that listeners are sensitive to the timing of the different acoustic information for the IP boundary as it becomes available in the auditory stimuli. Specifically, across this group of listeners, fixations on the image representing the IP boundary are predicted to increase when a boundary-adjacent syllable is lengthened relative to conditions without lengthening. Similarly, listeners’ fixations on the IP boundary image should increase when the stimuli contain a small or large pitch reset, while their fixations on the word boundary image should increase when there is no pitch reset in the auditory stimuli. Finally, a study by Zhang (2012) that used a similar design to examine cue-weighting of the primary IP boundary markers, but did not test the time course of perception, showed that listeners’ perception of a target boundary was influenced by pause duration significantly more than by final lengthening or pause. Therefore, the prediction is that the presence of a pause may override any potential influence of final lengthening on the time course of boundary perception.

More specific hypotheses and predictions for the perceptual consequences of the acoustic manipulations are presented in the results section for pause duration (3.3.1), pitch reset (3.3.2), and final lengthening (3.3.3).

3.3. Results

This section presents the results for the effects of different combinations of IP boundary cues on the perception of IP and word boundaries. Selection of GAMMs was done using the compareML function in R, by comparing the AIC scores between a full model and a corresponding, reduced model. The results of the model with the best fit are reported.

Participants' eye movements to the target images in each trial were monitored starting from the onset of the auditory stimulus and lasting through the end of the stimulus, but the time range included in all GAMMs reported in this chapter was from 200ms after TG1 onset to 400-500ms after TG1 offsets (which varied depending on the lengthening conditions). The only instance where a larger range is shown is for the pause duration, so as to show the general pattern of eye-movements.

All black vertical lines in the figures in this section indicate the temporal location of the TG1 offsets or the syllable boundaries of TG1 in the auditory stimuli, to which 200ms have been added. The 200ms addition is included given the estimate of around 200ms to program an eye movement upon hearing an auditory stimulus (e.g., Dahan et al., 2001). Thus, fixations beginning roughly 200ms after TG1 offsets might be associated with acoustic information that becomes available shortly after TG1 offsets. There are *multiple* vertical lines (i.e. TG1 offsets) because the durations of TG1 varied depending on the manipulations of syllable durations that were applied to the different patterns of final lengthening.

3.3.1. Presence vs. absence of pause

A greater proportion of fixations on the IP boundary image is predicted when listeners hear auditory stimuli with a pause between TG1 and TG2 than when they hear stimuli without that pause. Correspondingly, more fixations on the word boundary image are predicted when listeners hear auditory stimuli without a pause between TG1 and TG2.

The results, given in Figure 3.8, show that, as predicted, participants were sensitive to pause at the end of an IP. In nine out of 12 manipulation conditions (Figure 3.4), there was no pause between TG1 and TG2. On hearing these stimuli (left panel of figure), participants fixated on the image representing the word boundary (blue lines) but not the IP boundary (red lines). However, on hearing the remaining three manipulation conditions, which included a silent interval of 240ms between TG1 and TG2, participants fixated on the IP boundary image (red lines in the right panel) rather than the word boundary image (blue lines). The different patterns of responses show that presence of pause between pre- and post-boundary words is a salient cue for an IP boundary, and its absence indicates a word boundary. Given these robust effects, no statistical model was run to test the effect of the presence and absence of pause duration on the perception of prosodic boundary between TG1 and TG2.

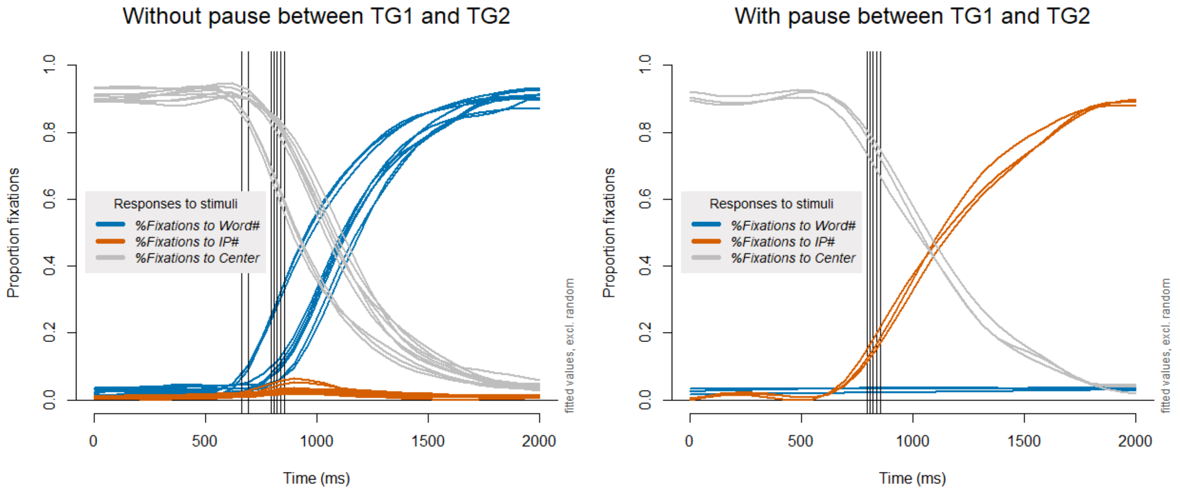


Figure 3.8. Proportion fixations on IP (red) or word (blue) boundary images in response to the 12 different types of auditory stimuli, without (left panel) or with (right) pause between TG1 and TG2. The grey lines represent proportion fixations on the center image, which listeners were directed to look at before hearing the auditory stimulus. The multiple vertical lines represent the TG1 offsets (plus 200ms to program an eye movement), which varied depending on the lengthening patterns. Time 0 = onset of TG1 in the auditory stimuli.

3.3.2. *Pitch reset*

This section presents the results of the models that tested for the influence of pitch reset between TG1 and TG2 on the perception of IP and word boundaries. The effects of pitch reset without pause, and without (3.3.2.1) or with (3.3.2.2) final lengthening in TG1 are presented. The effect of pitch reset could not be tested for the three manipulation conditions involving pause because the small pitch reset was present in all three conditions.

It is hypothesized that listeners will be sensitive to pitch reset. In the absence of final lengthening (or pause), this would mean fewer proportion fixations on the word boundary image when the stimuli contain small or large pitch reset than when they contain no pitch reset. However, when the stimuli contain both pitch reset and final lengthening, or all three cues including pause duration, these manipulations may interact or have an additive effect, possibly yielding even fewer fixations on the word boundary image when the stimuli contain pitch reset and final lengthening than when the stimuli contain pitch reset only. Moreover, if listeners are sensitive to the size of pitch reset, they should show fewer fixations on the word boundary image when they hear the large pitch reset than the small pitch reset.

3.3.2.1. Pitch reset without final lengthening

There are two hypotheses that concern pitch reset: one concerns the effect of the presence or absence of pitch reset (M0 vs. M1 in Figure 3.4) on the timing of fixations on the word boundary, while the other concerns the effect of the size of pitch reset (M1 vs. M2). These hypotheses are tested in a single model.

The GAMM included Reset Type (no reset vs. small reset vs. larger reset) as a predictor variable, the difference smooth that fitted the difference between each pair of the reset types, and Participant as a random smooth. The model tested participants' proportion fixations on the word boundary image for stimuli that did not contain final lengthening in TG1 or pause between TG1 and TG2 (M0, M1, M2).

The model structure and output are shown in Table 3.4. The parametric terms in the output indicate that the difference between proportion fixations on the word boundary image in response to the no vs. small reset stimuli is marginally significant (p=0.053), while the difference between the responses to the no vs. large reset stimuli is not significant (p=.18). The difference smooths, however, indicate that both differences are significant (p<0.001 for both), meaning that the shapes of the smooths in these comparisons are significantly different.

Model Structure:

Proportion fixation ~ Manipulation Type + s(Time) + s(Time, by=Manipulation Type) + s(Time, Participant)

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.473201	0.020326	23.281	<2e-16 ***
M1	0.004735	0.002452	1.931	0.0534 .
M2	0.003325	0.002452	1.356	0.1750

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Time)	8.549	8.678	41.375	<2e-16 ***
s(Time):M1	4.872	5.915	3.950	6e-04 ***
s(Time):M2	5.944	7.063	8.255	3.86e-10 ***
s(Time,subj)	155.189	170.000	73.413	< 2e-16 ***

Table 3.4. Structure and output of model testing for the effect of Reset Type.

To identify the time range over which the smooths significantly differ and how they differ, the three smooths are plotted, along with the three difference smooths in Figure 3.9.

Significant differences (i.e., regions that significantly differ from zero) are indicated in the intervals delineated in red along the x-axis.

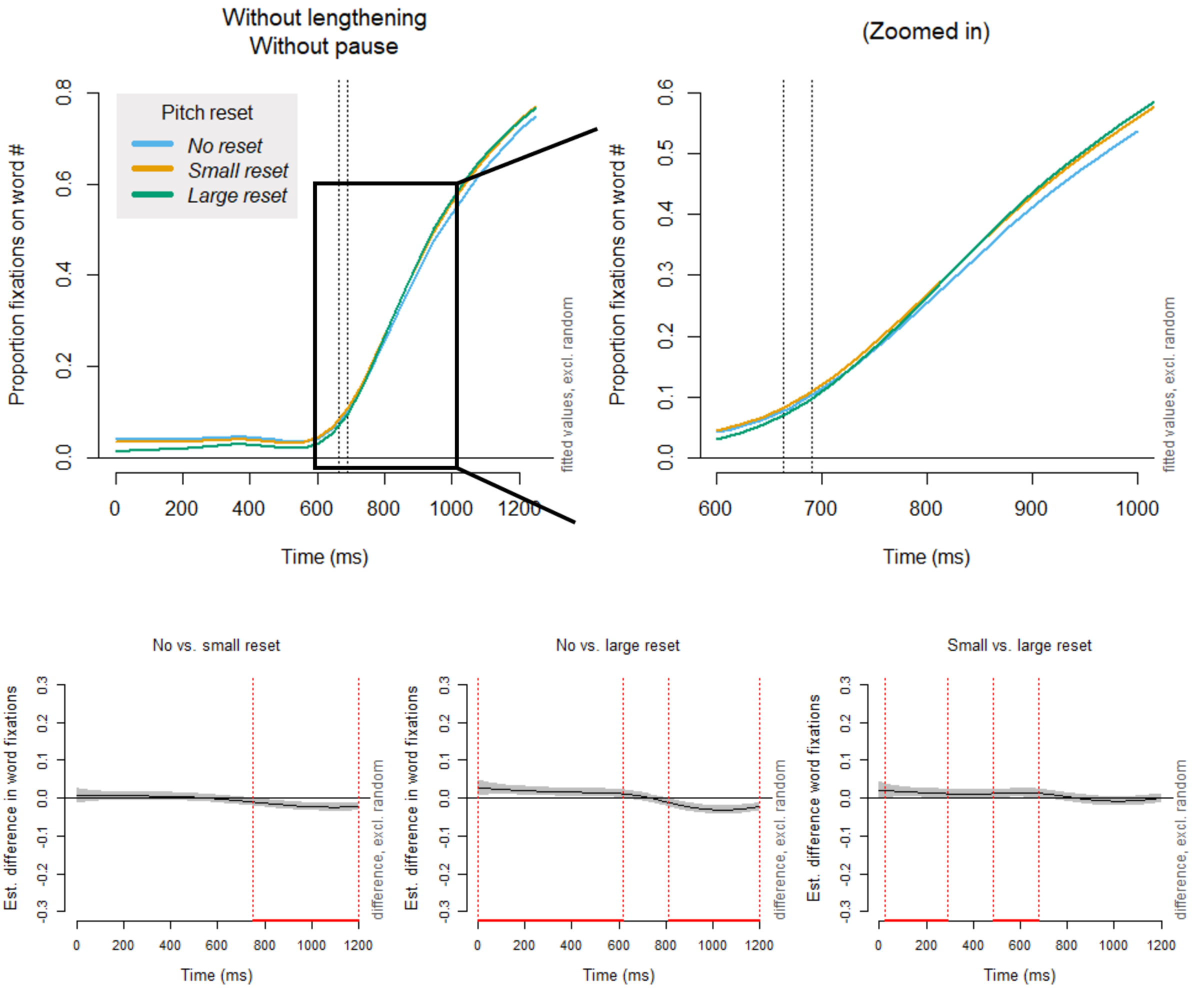


Figure 3.9. Model predictions for pitch reset results. Top left: Smooths fitted for the three Reset Types. Bottom panels: Difference smooths for each pair of Reset Types. Time 0 = onset of TG1 in auditory stimuli. Vertical dotted black lines = TG1 offsets plus 200ms for two TG1s (earlier for ‘naNIna’).

The top panel in Figure 3.9 shows model predictions for fixations on the word image in three Reset Types over time, while the bottom three panels show the difference between each pair of smooths, with the time ranges over which differences are significant being identified and marked in red. The model predictions in Figure 3.9 show a rapid increase in proportion fixations on the word boundary image in all three reset type conditions shortly after the adjusted (+200ms) TG1 offsets (marked by the vertical lines), indicating that listeners perceived a word boundary despite the f0 information that signals an IP boundary.

In testing for the effect of Reset Type, because TG1 offsets co-occur with pitch reset (or lack thereof), a significant difference in the timing of proportion fixations on target after the adjusted TG1 offset would suggest an effect of Reset Type. The three bottom panels of Figure 3.9 comparing the smooths for the no vs. small vs. large reset types show that there are significant differences in the no vs. small and no vs. large reset comparisons, but not in the small vs. larger reset comparison. For the first two comparisons, a significant difference occurring after adjusted TG1 offset (i.e., after about 700ms) would seem to be attributable to pitch reset, given that pitch reset is the only difference across the manipulations that occurs after TG1 (and continues throughout the span of TG2). Contrary to the expected positive differences between the pitch reset conditions, though, the model-predicted negative differences in the temporal region of interest between no vs. small (left panel) and large (middle panel) reset types indicate that listeners fixated (slightly but significantly) more often on the word boundary image when the stimuli included a reset. The lack of significant difference after adjusted TG1 offset between small vs. large reset (right panel) suggests that the difference in the size of pitch reset did not influence perceptual responses to the auditory stimuli.

Lastly, the middle and right panels show small but significant differences in smooths *before* TG1 offsets – differences that cannot be meaningfully linked to the variation in reset types, which occurs after TG1 offset. While the raw data are not shown in the dissertation, a careful inspection of these data showed that a few participants fixated on the target images (rather than the center image) earlier in the trial, prior to any crucial difference in the region. Given this, it may be speculated that these regularly occurring preemptive fixations may have contributed to the significant differences; Since this is a region with otherwise very little variability (due to few fixations on target images by the other participants), such small differences might have led to the model's prediction of significant differences.

3.3.2.2. Pitch reset with final lengthening

If the combination of pitch reset and final lengthening induces a stronger IP boundary percept than pitch reset or final lengthening alone, there should be fewer fixations on the word boundary image when the stimuli contain both cues than when the stimuli contain only one cue. However, because the critical time point (i.e., of when the last cue became available to listeners) varied depending on which cue(s) – pitch reset or final lengthening or both – is/are present solely or in conjunction in the auditory stimuli, the additive effect (pitch reset + final lengthening vs. pitch reset only vs. final lengthening only) was not tested in a statistical model.

Instead, in the models reported in this section, the effect of presence vs. absence of pitch reset on word boundary fixations was examined for manipulations in which the auditory stimuli contained three different patterns of final lengthening in TG1: early onset lengthening, late onset lengthening, and late onset lengthening + S1 shortening (see M3~M8 in Figure 3.4). Because all stimuli tested in these models involve some pattern of final lengthening, the effect of presence of pitch reset is comparable to the presence of the additive effects of pitch reset and final lengthening. Absence of pitch reset, though, involves conflicting boundary information: no pitch reset is information for a word boundary whereas lengthening is information for an IP boundary.

The structure and output of the models comparing proportion target fixations as a function of Reset Type for each of the three patterns of final lengthening are given in Tables 3.5 (early onset lengthening), 3.6 (late onset lengthening), and 3.7 (late onset lengthening + S1 shortening). For early onset lengthening (M3 vs. M4), the effect of Reset Type (no vs. small reset) is significant (p<.001), suggesting that there is an overall difference between proportion fixations for the two types of stimuli. The smooth term that modeled the difference between the shapes of smooths for the two conditions over time is significant (p<.001), meaning that there are significant differences over time between stimuli with no vs. small reset, in the presence of early onset lengthening. That is, listeners fixated more often on the word boundary image when the stimuli contained no pitch reset than when it contained a small reset, which is in line with the prediction for the effect of pitch reset.

For both late onset lengthening (M5 vs. M6) and late onset lengthening + S1 shortening (M7 vs. M8), the effect of Reset Type is significant (p<.05 for late onset lengthening; p<.001 for late onset lengthening + S1 shortening), and the difference smooth term is also significant (p<.05 for late onset lengthening; p<.01 for late onset lengthening + S1 shortening).

Model Structure:				
Proportion fixation ~ Manipulation Type + s(Time) + s(Time, by=Manipulation Type) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.368189	0.020826	17.679	< 2e-16 ***
M4	0.010110	0.002384	4.241	2.23e-05 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	8.29	8.496	36.212	< 2e-16 ***
s(Time):M4	6.46	7.572	7.625	7.05e-10 ***
s(Time,subj)	152.28	170.00	44.758	< 2e-16 ***

Table 3.5. Structure and output of the model for testing the effect of Reset Type when S2 and S3 of TG1 are lengthened (i.e., early onset of final lengthening).

Model Structure:				
Proportion fixation ~ Manipulation Type + s(Time) + s(Time, by=Manipulation Type) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.404759	0.021068	19.212	< 2e-16 ***
M6	-0.008775	0.002350	-3.734	0.000189 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	8.291	8.490	36.516	< 2e-16 ***
s(Time):M6	5.089	6.169	3.393	0.00212 **
s(Time,subj)	152.942	170.000	62.263	< 2e-16 ***

Table 3.6. Structure and output of the model for testing the effect of Reset Type when S3 is lengthened (i.e., late onset of lengthening).

Model Structure:				
Proportion fixation ~ Manipulation Type + s(Time) + s(Time, by=Manipulation Type) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.393362	0.018083	21.753	< 2e-16 ***
M8	0.012971	0.002494	5.201	1.99e-07 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	8.332	8.535	46.376	< 2e-16 ***
s(Time):M8	5.151	6.235	5.178	1.53e-05 ***
s(Time,subj)	146.615	170.00	37.695	< 2e-16 ***

Table 3.7. Structure and output of the model for testing the effect of Reset Type when S1 is shortened and S3 is lengthened (i.e., late onset of lengthening + S1 shortened).

The model predictions for the effect of Reset Type (no vs. small reset) in the three different patterns of final lengthening as well as the difference smooth plots with the significant time range (marked in red) are shown in Figure 3.10. As with the previous model for the effect of Reset Types (3.3.2.1), the significant differences identified in the difference plots prior to TG1 offsets (marked with vertical lines in the model predictions) cannot be meaningfully linked to the effect of reset types, which is associated with the time range approximately 200ms after TG1 offsets in the auditory stimuli.

The effect of reset types after the pattern-specific adjusted TG1 offset is shown to be significant in early onset lengthening (top panel in Figure 3.10) and in late onset lengthening + S1 shortening (bottom panel), but not in late onset lengthening (middle panel). For the early onset lengthening condition, after adjusted TG1 offset (indicated by the vertical lines), listeners were more likely to fixate on the word boundary image when the stimuli had no pitch reset than small reset, as predicted, consistent with absence of pitch reset contributing to the percept of a word boundary in the presence of conflicting early lengthening. On the other hand, for the late onset lengthening condition, where there is presumably less conflict in the acoustic information for boundaries, there is no significant difference after adjusted TG1 offset. For the late onset lengthening + S1 shortening condition, contrary to expectations, listeners were less likely to fixate on the word boundary image after TG1 offset when the stimuli had no reset than when they had small reset. In this case, it appears that the model predicts that the significant difference in responses to these manipulations induced by differences earlier in the stimuli (prior to TG1 offset) extends into the post-TG1 offset region. While the effect of the time period before TG1 offset cannot be related to the examined manipulations, the effect observed after TG1 offset (i.e., after the pitch reset manipulations) might be a carry-over effect.

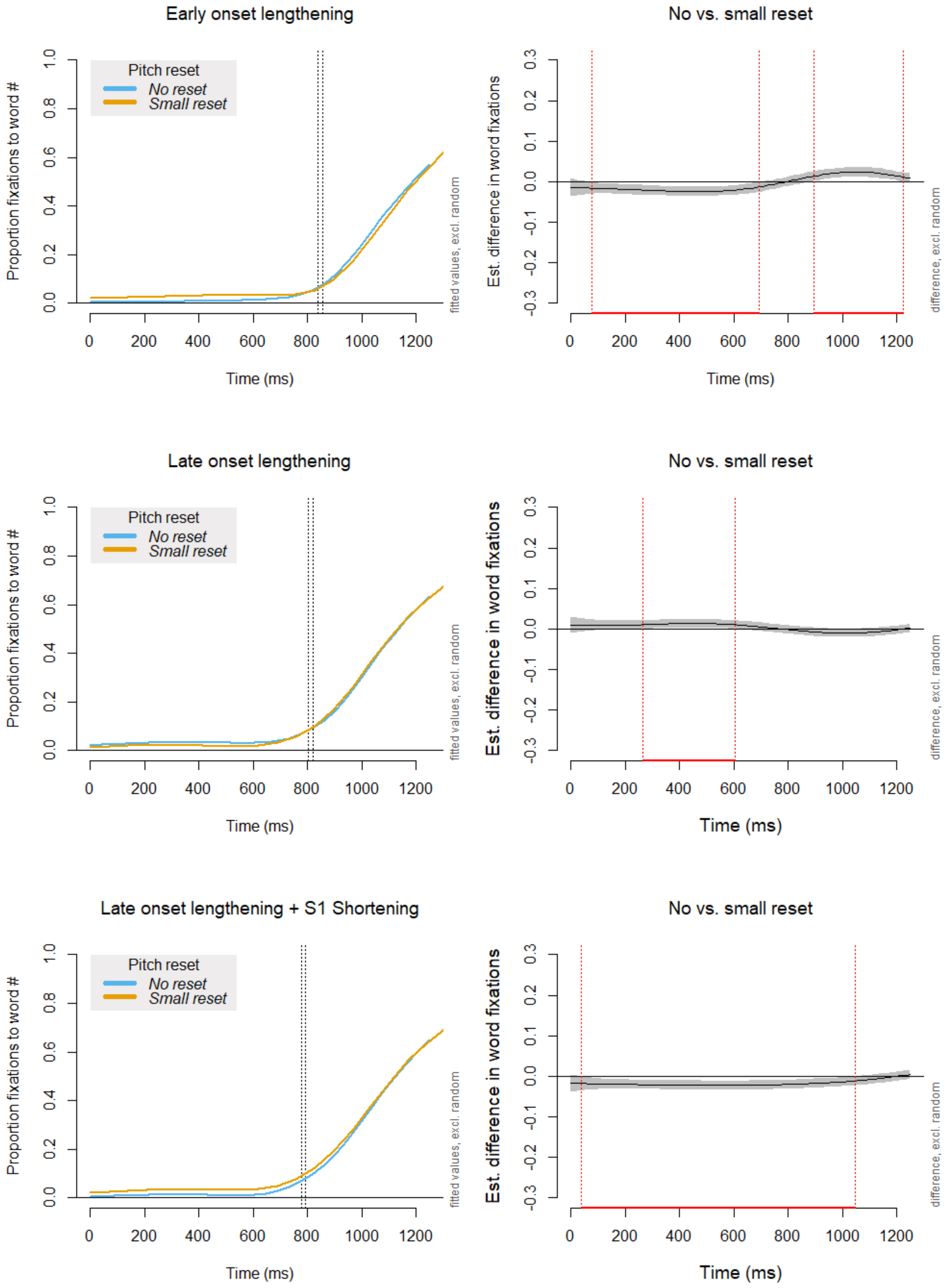


Figure 3.10. Model predictions for pitch reset results for three TG1 syllable duration patterns. Time 0 = TG1 onset + 200ms in auditory stimuli. Left panels: Proportion fixations on word boundary image over time. (Vertical lines: adjusted (+200ms) TG1 offset for two TG1s (earlier for 'naNIna'). Right panels: Difference smooths for each comparison. Grey shading: 95% CI.

3.3.3. *Effect of phrase-final lengthening*

This section presents the results of the models that tested for the influence of the three final lengthening patterns in TG1 on the perception of IP and word boundaries. The effects of the lengthening patterns without pause, and without (3.3.3.1) or with (3.3.3.2) pitch reset between TG1 and TG2 are presented, as are the effects with both pitch reset and pause (3.3.3.3).

Recall that three manipulation methods were applied to the auditory stimuli, which corresponded to the three primary acoustic characteristics associated with perception of the IP boundary. Figures 3.11 and 3.12 show the relative timing of the available temporal information across the lengthening conditions for 'maMIma' and 'naNIna' trials, respectively.

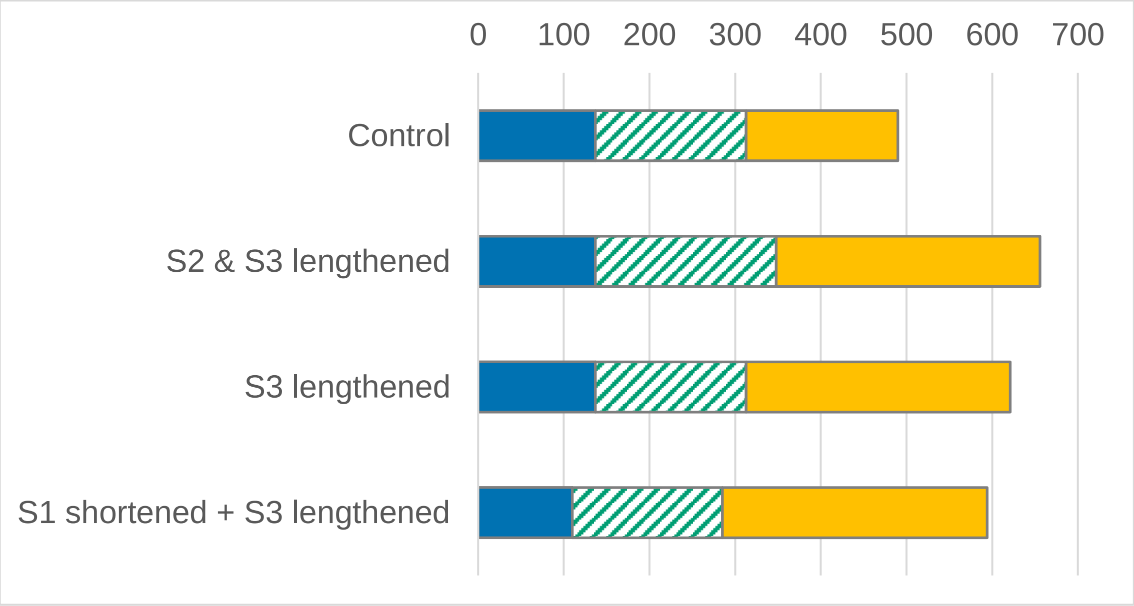


Figure 3.11. Time points of syllable onset/offset of TG1 ('maMIma') in three patterns of phrase-final lengthening and the Control condition. [Blue: S1; Green: S2; Yellow: S3]

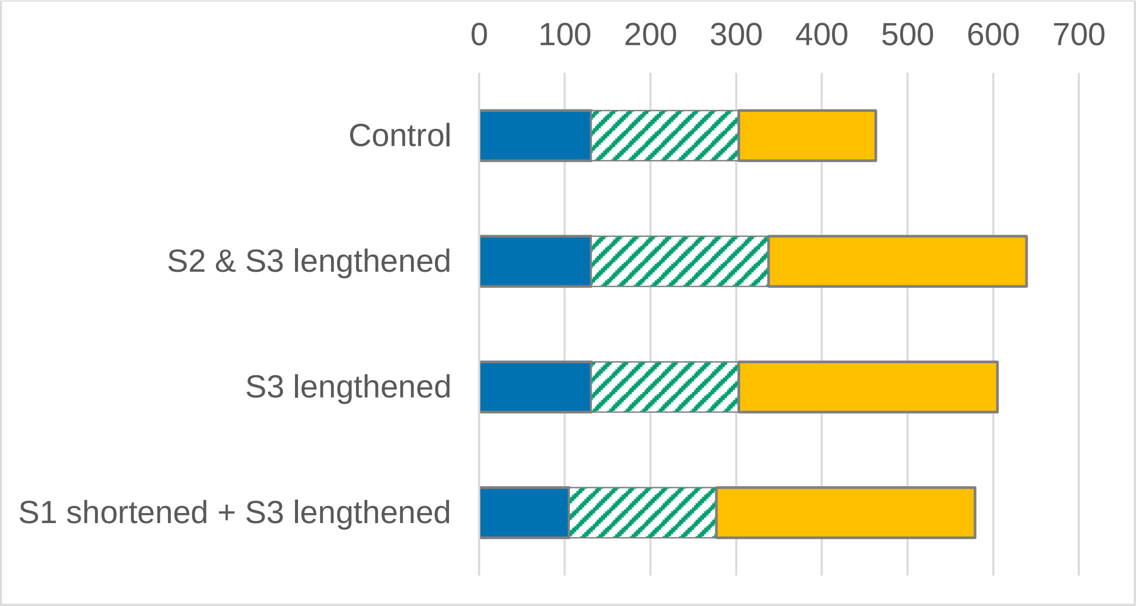


Figure 3.12. Time points of syllable onset/offset of TG1 (‘naNIna’) in three patterns of phrase-final lengthening and the Control condition. [Blue: S1; Green: S2; Yellow: S3]

The most common pattern observed in the production experiment was lengthening of S2 and S3 of TG1 (“early onset of lengthening”), the second most common pattern was lengthening of just S3 (“late onset of lengthening)”, and the third was S3 lengthening co-occurring with S1 shortening (“late onset of lengthening + S1 shortening”).

Unlike the effect of Reset Type, in which stimuli have the same syllable durations within comparisons, the time at which the final lengthening cue becomes available varies. Final lengthening occurs the earliest in early onset lengthening condition, and latest in late onset lengthening condition. However, if listeners interpret the shortening in the late onset lengthening +S1 shortening condition as a cue to an upcoming final lengthening (and therefore the boundary), listeners will get boundary information in the following order (see onset/offset of S3 in Figure 3.12): late onset lengthening + S1 shortening (earliest), late onset lengthening, early onset lengthening (latest). Because listeners get durational information about the upcoming boundary at different time points, it would be misleading to compare proportion fixations for stimuli with different lengthening patterns for the same time range. Therefore, the effect of lengthening conditions is examined by comparing proportion fixations on the word boundary image at different time points, such as the syllable boundaries, which are specific to each condition.

The predictions for the effect of lengthening conditions are as follows: listeners are predicted to be sensitive to final lengthening, which is information for an upcoming IP boundary. In the absence of pitch reset or pause, final lengthening serves as conflicting information, in sequence: final lengthening in TG1 signals the IP boundary but onset of post-boundary TG2 with no pause or pitch reset signals a word boundary. However, if listeners use lengthening as information about an IP boundary even in the absence of other IP cues, then the earlier the information for an upcoming IP boundary, the less listeners should fixate on the word boundary image. This general prediction yields syllable-specific predictions for S1, S2, and S3.

There are two possibilities: First, suppose listeners are not sensitive to the shortening boundary effects (on S1 in conditions M7 and M8) but are sensitive to final lengthening. In this case, proportion fixations for all three lengthening conditions at S1 offset will be comparable to those for the control condition with no lengthening. At S2 offset, fewer proportion fixations on the word boundary image would manifest for the early onset lengthening condition but not in the other conditions. At S3 offset, the control condition would show more proportion fixations on the word boundary image than the other conditions, while the early onset lengthening condition would show fewer proportion fixations. After TG1 offset, listeners will fixate on the word boundary image due to the absence of pause (and instead the presence of the TG2 onset that follows TG1 offset).

Alternatively, if listeners are sensitive to syllable shortening, which they interpret as a cue for upcoming final lengthening, they would fixate least often on the word boundary image in the late onset lengthening + S1 shortening conditions.

The main prediction for the perceptual effect of lengthening in manipulations with pitch reset and no pause (M3~M8) – that is, stimuli in which the intonational cue becomes available immediately after the temporal information – is as follows. Relative to the corresponding manipulations without the intonational information, presence of (small) pitch reset may lead to further reduction in fixations on the word boundary image.

Lastly, when all three IP boundary cues are present in the target word sequence (final lengthening, pitch reset, and pause; M9~M11), listeners hear a pause after TG1 instead of TG2. Because pause is salient information for an IP boundary (section 3.3.1), proportion fixations on the IP, rather than word, boundary image will be examined to test the hypotheses. For these stimuli, the order in which listeners hear these cues is final lengthening > pause duration > pitch reset. The predictions regarding the differences in proportion fixations at syllable boundaries between the lengthening conditions can be applied here, up until TG1 offset, but in the opposite direction because proportion fixations to the IP boundary image are being calculated. After TG1 offset, listeners should fixate on the IP boundary image as a pause becomes available. Comparing the effect of the lengthening conditions using stimuli with or without pause, but with the other two cues present, allows us to examine how differences between proportion fixations at given syllable boundaries are related to (1) which cue is followed by the final lengthening cue (pitch reset vs. pause), and (2) which boundary listeners eventually perceive (IP vs. word boundary).

3.3.3.1. Phrase-final lengthening without pitch reset or pause

To examine whether auditory stimuli that differ only in phrase-final lengthening patterns elicit different patterns of fixations, fixations on the word boundary image are modeled in a GAMM that included Lengthening Pattern as a predictor variable, the difference smooth that fitted the difference between each pair of the patterns of final lengthening, and Participant as a random smooth. The model structure and output are given in Table 3.8. The effect of lengthening pattern is significant, such that proportion fixations on the word boundary image between all four conditions are significantly different. The differences in the changes in proportion fixations over time, shown in significance of smooth terms, are also significant.

Model Structure:				
Proportion fixation ~ Manipulation Type + s(Time) + s(Time, by=Manipulation Type) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.467730	0.016845	27.77	<2e-16 ***
M3	-0.093334	0.002449	-38.11	<2e-16 ***
M5	-0.063670	0.002449	-26.00	<2e-16 ***
M7	-0.069762	0.002449	-28.48	<2e-16 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	8.439	8.667	56.58	<2e-16 ***
s(Time):M3	8.727	8.960	157.54	<2e-16 ***
s(Time):M5	8.640	8.941	126.60	<2e-16 ***
s(Time):M7	8.523	8.909	78.84	<2e-16 ***
s(Time,subj)	154.460	170.00	75.69	<2e-16 ***

Table 3.8. Structure and output of the model for testing for the effect of Lengthening Pattern (without pitch reset or pause).

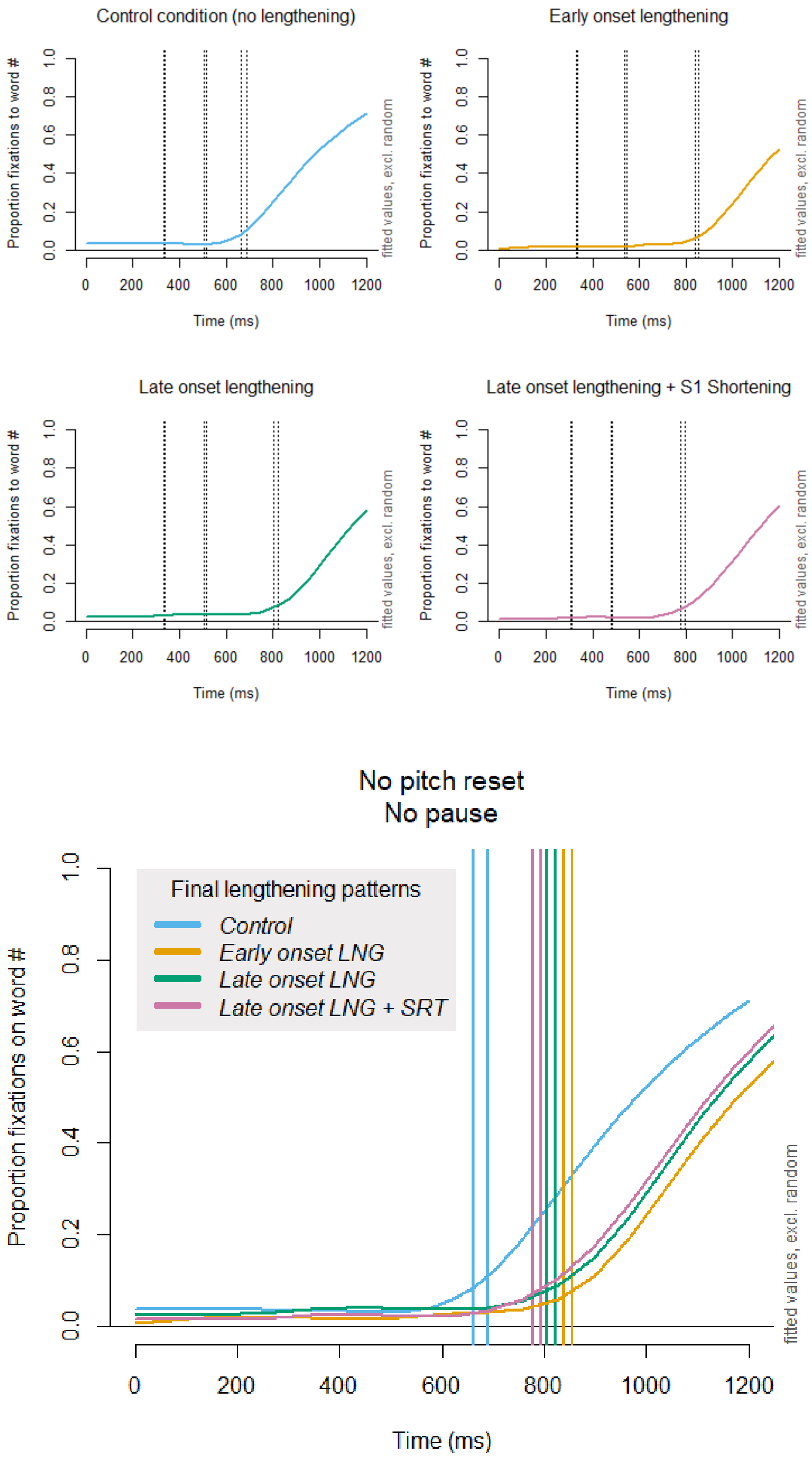


Figure 3.14. Model predictions of proportion fixations on the word boundary image for the three different lengthening patterns and the control condition. The pairs of black vertical lines in the four individual panels are syllable boundaries of TG1, while the pairs of color-coded vertical lines in the last panel are TG1 offsets for different lengthening patterns (with the 200ms programming lags). In all pairs of vertical lines, the earlier one is for TG1 ‘naNIna’. Time 0 = onset of TG1 in auditory stimuli.

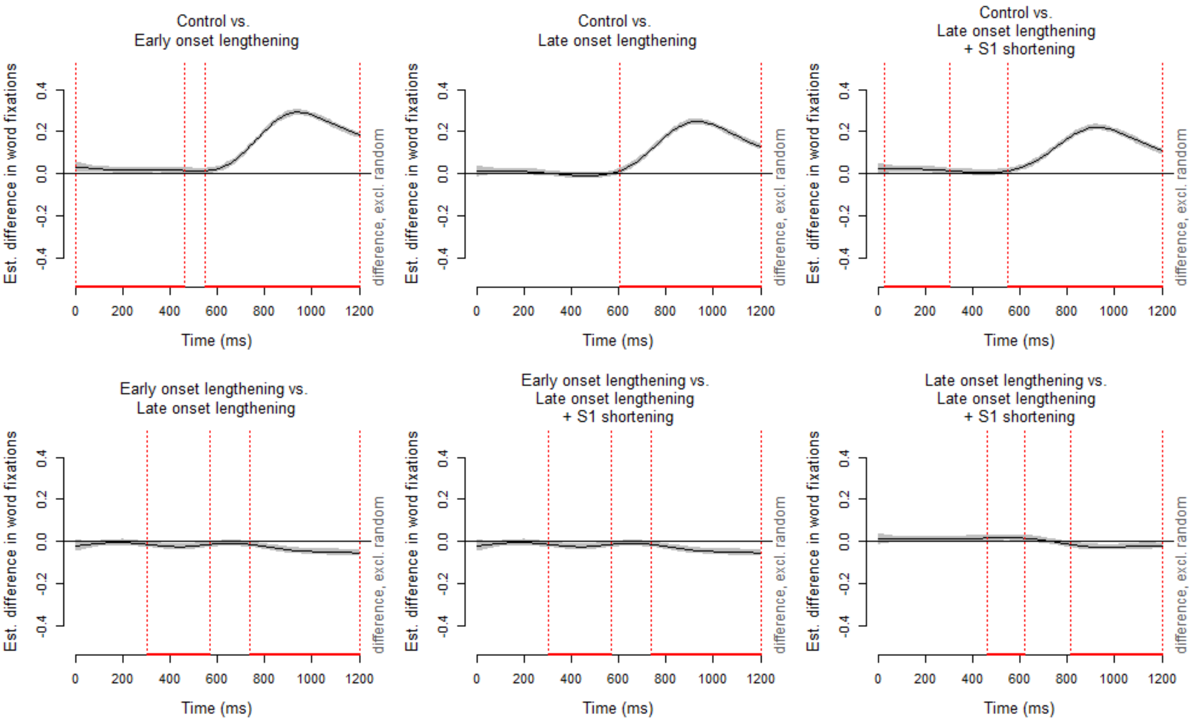


Figure 3.15. Difference smooths for each comparison of final lengthening manipulations. Grey shading: 95% CI.

Visual inspection of the model predictions for different lengthening patterns offers more detailed analysis of the results. The model predictions shown in Figure 3.14 illustrate proportion fixations on the word boundary image over time in relation to the color-coded TG1 offsets (with the 200ms programming lag added). In all four conditions, the model predicts that fixations on the word boundary image start rapidly increasing after adjusted TG1 offset. For example, for the control condition (blue) with no lengthening, listeners are predicted to fixate the word boundary image about 10% of the time shortly after TG1 offsets (approx. 650-700ms) and about 50% of the time at 1000ms. For the lengthening conditions, word boundary image fixations increase later than for the control condition, and timing of this increase varies across the three lengthening conditions. Moreover, the influence of lengthening on target fixations extends throughout TG2 and beyond – that is, well past the word boundary information signaled by the lack of pause or pitch reset. The difference smooths shown in Figure 3.15 indicate that these differences are significant, suggesting that listeners are sensitive to the different lengthening manipulations that resulted in the different timings of TG1 offsets.

The difference smooths comparing the no-lengthening control condition to the three lengthening conditions (top panels of Figure 3.15) show the expected positive – and large – differences, consistent with listeners looking earlier and more often to the word boundary image when no lengthening is present. Of course, this is to be expected in part because TG2 starts earlier in the control condition, but it is noteworthy that the significant differences continue up through 1200ms into the trial. The difference smooths comparing the lengthening conditions (bottom panels) also show the expected – in this case, negative – differences: the later the offset of TG1 (pink, green, and yellow vertical lines in Figure 3.14), the fewer and later the looks to the word boundary image, indicating (i) that listeners begin looking at the word boundary image nearly as soon as clear information about that boundary (no pause and no pitch reset) becomes available but (ii) the effects of (conflicting) lengthening extend well into the post-boundary (TG2) portion of the trials.

Proportion fixations at S1 and S2 boundaries of TG1 (i.e., in the adjusted roughly 400-600ms region) are also predicted to differ if listeners are sensitive to the precise details of that temporal information. For early vs. late onset lengthening (left bottom panel of Figure 3.15), results are as expected: there are fewer predicted proportion fixations on the word boundary image in the early than the late onset condition, as indicated by the negative difference in the 400-600ms region. However, listeners do not appear to be sensitive to shortening during TG1 given that the difference smooth for early onset lengthening vs. late onset +S1 shortening (middle panel) is essentially the same as for the same comparison without shortening (left panel). Moreover, comparison of the late onset lengthening conditions that differ in presence/absence of S1 shortening (right panel) shows a very small but significant effect during TG1 in the unexpected direction.

3.3.3.2. Phrase-final lengthening with pitch reset and no pause

A GAMM with the same structure as that used in 3.3.3.1 tested the effects of lengthening pattern on listeners' responses to auditory stimuli that included no pause and a small pitch reset between TG1 and TG2. The model structure and output are given in Table 3.9. The parametric terms indicate significant effects of the different lengthening conditions on fixations on the word boundary image. The significance for the smooth terms suggests that fixations over time significantly differed between the different lengthening conditions.

Model Structure:				
Proportion fixation ~ Manipulation Type + s(Time) + s(Time, by=Manipulation Type) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.379091	0.018936	20.020	< 2e-16 ***
M6	0.010779	0.002448	4.404	
M8	0.026434	0.002448	10.799	< 2e-16 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	8.299	8.510	43.28	< 2e-16 ***
s(Time):M6	7.421	8.330	29.06	< 2e-16 ***
s(Time):M8	7.661	8.489	37.63	< 2e-16 ***
s(Time,subj)	153.664	170.000	61.51	< 2e-16 ***

Table 3.9. Structure and output of the model for testing the effect of Lengthening Pattern (with pitch reset and no pause).

Figure 3.16 visualizes the model predictions for the three lengthening conditions when a small pitch reset was present. Note that the three lengthening patterns compared in this model are the same as in the previous model (minus the control condition); the only difference is that the current model estimates responses to stimuli with a small pitch reset across the target boundary.

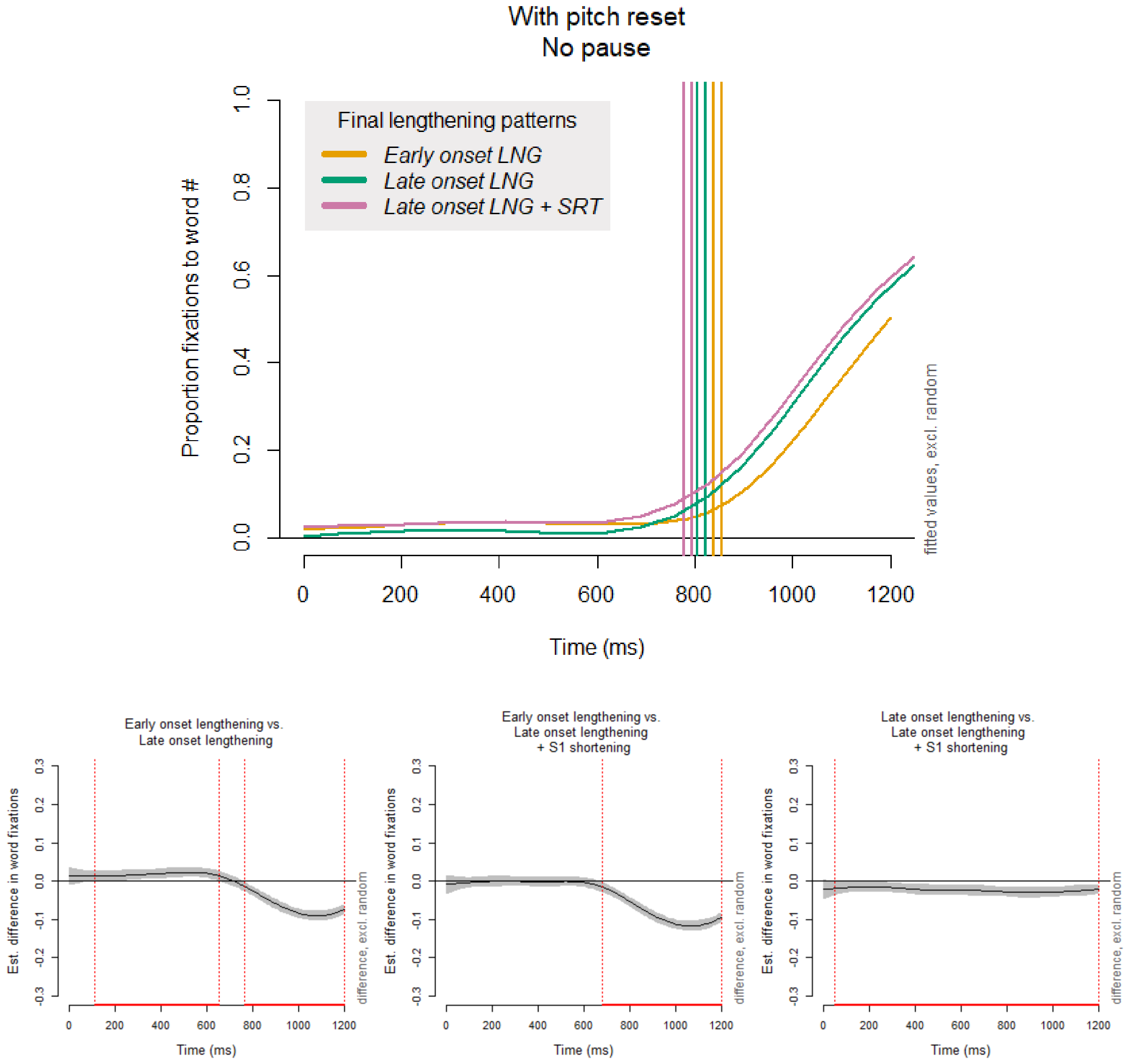


Figure 3.16. Model predictions of proportion word boundary fixations for three lengthening conditions with pitch reset. The pairs of color-coded vertical lines in the last panel are adjusted (+200ms) TG1 offsets for different lengthening patterns. In all pairs of vertical lines, the earlier one is for TG1 ‘naNIna’. Time 0 = onset of TG1 in auditory stimuli. Bottom three panels: difference smooths for each comparison of final lengthening manipulations. Grey shading: 95% CI.

The model predictions in Figure 3.16 indicate that the earlier the offset of TG1/onset of TG2, the earlier listeners fixate on the word boundary image, replicating the general pattern observed in the previous comparison. The significant differences at TG1 offset indicated in the difference smooth plots suggest that listeners are sensitive to the TG1 offset/TG2 onset and interpret the lack of pause at TG1 offset as a cue for the word boundary.

3.3.3.3. Phrase-final lengthening with both pitch reset and pause

To test whether variation in final lengthening influences perception of the target boundary when both additional IP boundary cues (i.e., pitch reset and pause) are present, perceptual responses to stimuli containing all three cues were compared in a GAMM with the same structure, with the exception that the response variable was fixations on the IP boundary image rather than the word boundary image.

It is predicted that, if listeners interpret the shortening in S1 as a cue for final lengthening, they will fixate more on the IP boundary image when the stimuli have S1 shortening than when the stimuli do not. In addition, if listeners are sensitive to the different timing of lengthening in the auditory stimuli, there will be a significant difference in proportion fixations on the IP boundary image at TG1 offset: listeners would fixate more often on the IP boundary image in the early onset lengthening condition than the late onset lengthening condition.

The model structure and output are given in Table 3.10. Both parametric terms are significant, indicating that there are significant differences between the three conditions of lengthening patterns. The difference smooth terms are also significant, suggesting that the shape of the response trajectories for the lengthening conditions are significantly different.

Model Structure:				
Proportion fixation ~ Manipulation Type + s(Time) + s(Time, by=Manipulation Type) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.399409	0.021749	18.364	< 2e-16 ***
M10	-0.009533	0.002407	-3.961	7.46e-05 ***
M11	-0.018899	0.002405	-7.860	3.89e-15 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	8.043	8.295	31.52	<2e-16 ***
s(Time):M10	7.187	8.159	13.44	<2e-16 ***
s(Time):M11	6.348	7.439	14.15	<2e-16 ***
s(Time,subj)	157.049	170.00	124.62	<2e-16 ***

Table 3.10. Structure and output of model testing for the effect of Lengthening Pattern.

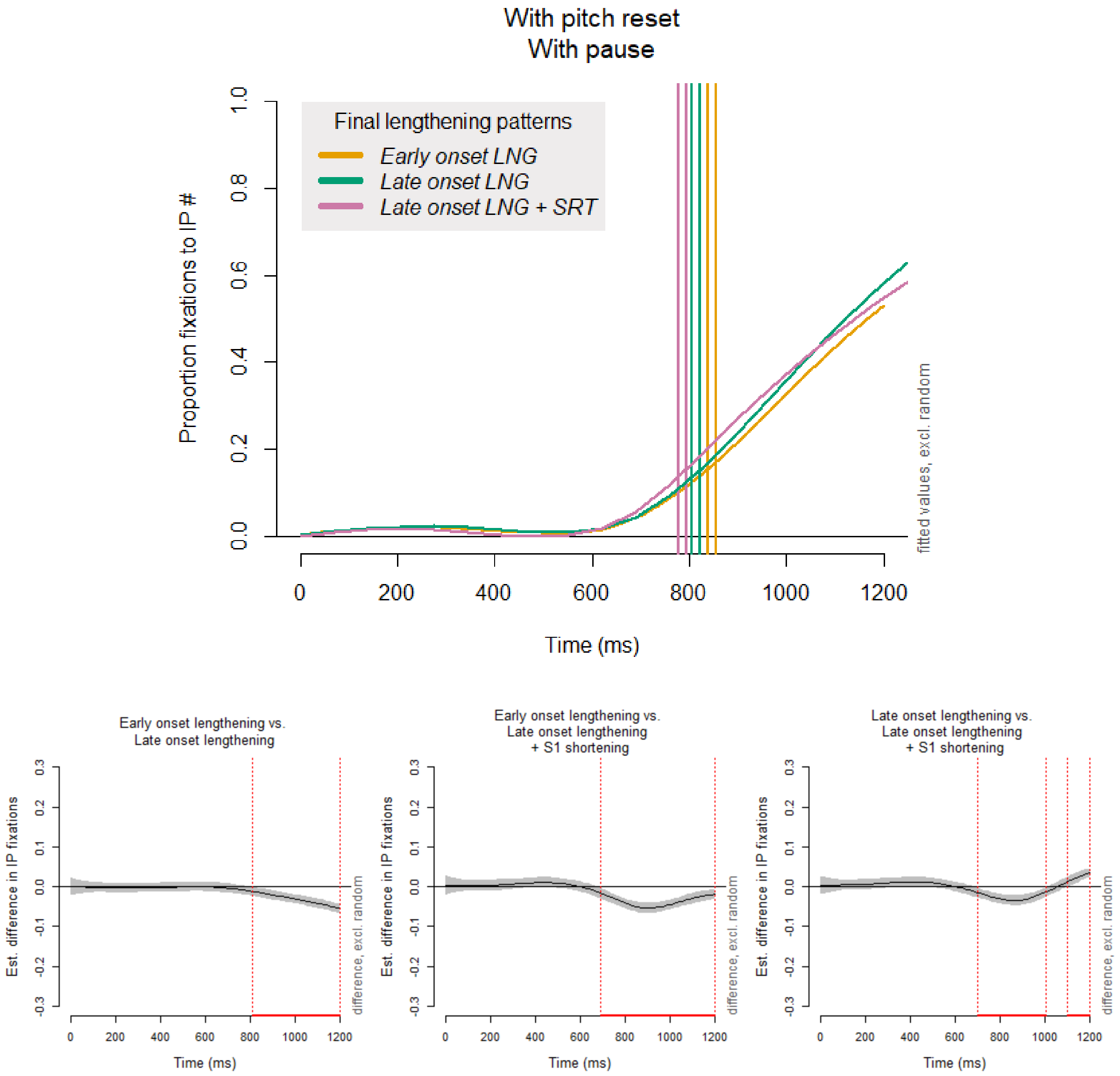


Figure 3.17. Model predictions of proportion fixations for three lengthening conditions with pitch reset and pause. The pairs of color-coded vertical lines in the last panel are adjusted (+200ms) TG1 offsets for different lengthening patterns. In all pairs of vertical lines, the earlier one is for TG1 ‘naNIna’. Time 0 = onset of TG1 in auditory stimuli. Bottom three panels: difference smooths for each comparison of final lengthening manipulations. Grey shading: 95% CI.

Figure 3.17 shows model predictions for the three lengthening patterns, allowing visual inspection of the timing of fixations on the IP boundary image in relation to the different IP markers. The order in which proportion fixations on the IP boundary image increase appears to match the order of the (adjusted) TG1 offsets of the lengthening conditions. The difference smooths indicate that these differences are significant, suggesting that listeners perceived an IP boundary at TG1 offset. Although the negative difference between the early vs. late onset lengthening conditions is in the unexpected direction (left bottom panel), the negative differences for the other two comparisons are consistent with listeners looking earlier to the IP boundary image when that boundary information includes S1 shortening.

3.4. Discussion of the perceptual results

The analyses of the eye tracking data showed that the acoustic manipulations of the three cues to IP boundary resulted in systematic differences in listeners' perceptual responses. Listeners are sensitive to pause duration and to some extent to final lengthening, but apparently not to pitch reset. The following subsections discuss perception of each IP boundary cue and how it relates to the perception of prosodic structure in general.

3.4.1. Perception of Pause Duration

The robust effect of pause duration in the perception of IP boundaries reported in the current study adds support to previous research which showed that pause is a salient cue to IP boundaries. When a pause was present, participants fixated on the IP boundary image, whereas they fixated on the word boundary image when there was no pause between TG1 and TG2 (Figure 3.8).

Pause duration seemed to outweigh final lengthening and pitch reset in the perception of boundaries, given the finding that listeners consistently fixated on the word boundary image when there was no pause even when both final lengthening and pitch reset were present. This result is in line with Petrone et al. (2017), who found that pause was perceived more categorically than f0 or final lengthening by German-speaking listeners. It should be noted, however, that the current study did not test pauses of different durations, but only the presence and absence of pause. It remains an empirical question whether listeners' responses to pauses of different durations would show more variation (that is, whether variable pause durations would give rise to less categorical responses).

3.4.2. Perception of Pitch Reset

Results concerning the effect of pitch reset were inconsistent and did not provide evidence that, as a group, listeners in this study relied on pitch reset in the perception of prosodic boundaries. Listeners fixated on the word boundary image more when the stimuli had no pitch reset than when it had small reset, but this pattern was found only in the early lengthening condition. The effect was absent in the late onset lengthening condition, and inconsistent with the prediction in the control condition (no lengthening) as well as in the late onset lengthening + S1 shortening condition. The different sizes of pitch reset did not yield significant differences in fixations. It is possible that the size of pitch reset derived from the model speaker's production may not have been salient for the listeners.

3.4.3. *Perception of Final Lengthening*

Participants fixated less on the word boundary image when the stimuli included final lengthening (of any pattern or magnitude) than when the stimuli did not, suggesting that the presence of final lengthening reduced the likelihood of perceiving a word boundary (Figure 3.14). This finding corroborates previous literature on final lengthening as a salient cue to the perception of IP boundaries. However, the results provide only limited evidence that listeners closely track temporal information during the pre-boundary word: during TG1, they appear to be sensitive to the presence of early lengthening but not to the presence of S1 shortening. In addition, listeners consistently fixated on the word boundary image when the auditory stimuli presented different patterns of final lengthening, alone and in conjunction with pitch reset. This finding suggests that final lengthening, at least in the magnitude used in this study, is not sufficient to induce a strong IP boundary percept.

For the hypothesis that shortening of S1 is a cue for upcoming final lengthening to be supported, the lengthening condition with S1 shortening should induce significantly fewer fixations on the word boundary image shortly after (adjusted) S1 offset. While such a pattern was not observed early in TG1, a pattern observed later in the trial, after TG1 offset, was that listeners fixated more on the IP boundary image when shortening was present.

3.4.4. *Perception of prosodic structure*

The results of the perception experiment showed that listeners are more sensitive to pause duration than to final lengthening or pitch reset. This finding contributes to the literature on cue weighting for the perception of prosodic boundaries by providing evidence that listeners may rely more on the presence/absence of pause than the other IP boundary markers. However, the results did not provide clear evidence that pitch reset and final lengthening were additive in their effects on boundary perception. For example, the timing of listeners' target fixations in response to stimuli containing final lengthening and pitch reset were not significantly different from their target fixations for stimuli containing only final lengthening (Figures 3.15 and 3.16). This is in line with previous research on cue-weighting in the perception of prosodic boundary (Yang et al., 2014; Gollrad et al., 2010).

CHAPTER 4

Individual Differences in the Production and Perception of Prosodic Boundaries

Chapter 3 investigated the perception of IP boundaries by testing whether listeners distinguish the auditory stimuli containing different combinations of the manipulated properties for IP boundaries. The results showed that listeners as a group are sensitive to the presence/absence of a pause as well as to the effects of final lengthening, while the findings for pitch reset were inconclusive. Given that the study presented in Chapter 2 found substantial individual differences in the production of IP boundaries, it is reasonable to assume that individuals also differed in the use of the information in the perception of prosodic boundaries. For example, Figure 4.1 illustrates, for two individual listeners, fixations over time on the target image relative to the syllable boundaries of TG1: Participant 9, with an average pause duration of 460ms (75th percentile), fixated on the IP boundary image earlier than Participant 11 with a much shorter average pause duration (235ms, 25th percentile). This chapter investigates whether and how the production and perception of the three acoustic properties of IP boundaries are related to each other at the level of the individual speaker-listener.

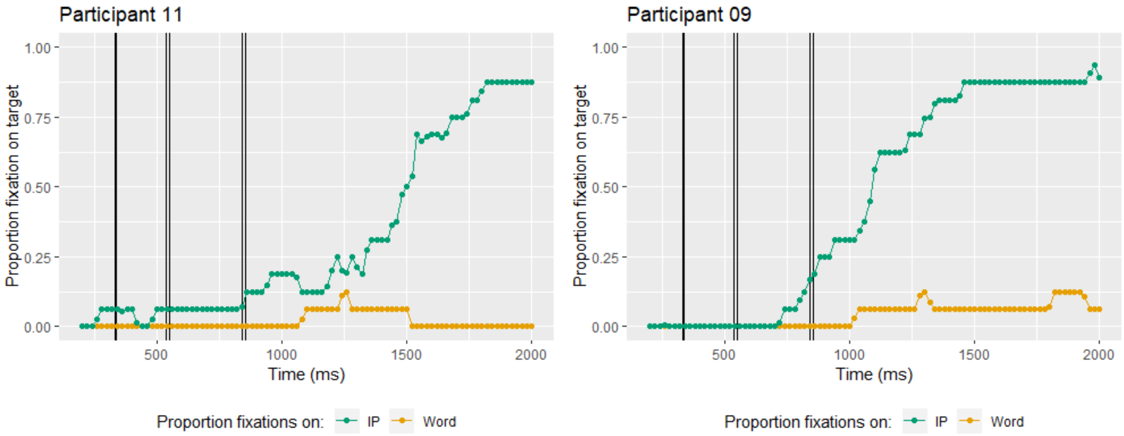


Figure 4.1. Proportion fixations for stimuli with early onset final lengthening, small pitch reset, and pause between TG1 and TG2. Participant 11 produced an average of 235ms of pause duration, and Participant 9' average was 460ms. The three pairs of vertical lines represent the syllable boundaries of TG1 (earlier for 'naNIna').

The relation between the production and perception of prosodic boundaries is examined in a set of Generalized Additive Mixed Model (GAMM) for each of the three acoustic properties for an IP boundary. The individual average values of the IP boundary markers, measured and analyzed in the production study, are included as continuous or categorical predictor variables in the combined models of production and perception. These models test whether the speaker-specific production data inform the perception model in a significant way.

The general hypothesis for the relation between production and perception of prosodic boundaries is that participants' production of a targeted property to differentiate IP from word boundaries will be reflected in their use of that property as information for an upcoming IP boundary. Section 4.1 presents the specific predictions for each boundary marker. Section 4.2 summarizes the results of the GAMM analyses for the relation between the production and perception of the three IP markers. Section 4.3 discusses the interpretation of the results. Section 4.4 summarizes the chapter and discusses limitations of the study.

4.1. Predictions for each IP boundary property

The statistical models reported in this chapter test the relation between the production results reported in Chapter 2 and the perception results reported in Chapter 3 for the 19 participants who completed both experiments (and who had reliable eye-tracking results). The perception results are, again, analyses of participants' eye movements as they listened to auditory stimuli that contained 12 different combinations of the three acoustic properties for IP boundaries.

Under the general hypothesis that production of the properties is reflected in perception, the rate at which proportion fixations on a target boundary increase over time is predicted to depend on speaker-specific production patterns.

For pause duration, individuals who produced longer pause on average should be more likely to attend to the pause cue, and thus respond to stimuli containing a pause between TG1 and TG2 earlier than those who produced shorter average pause. Similarly, individuals who produced larger f0 differences across IP boundaries should be more likely to be more sensitive to f0 information at a target boundary than those who produced smaller f0 differences. This means that, when the auditory stimuli present none of the acoustic properties that signal IP boundaries, lack of pitch reset should facilitate the response of the former individuals. Conversely, when there is small or large pitch reset in the absence of final lengthening or pause at the target boundary, there is conflicting information about the upcoming boundary, and individuals who attend to the f0 information should be slower to respond compared to those who do not use pitch reset in their production and, by hypothesis, do not find pitch reset to be as informative.

Lastly, in Chapter 2, individuals were divided into six groups depending on the syllable-wise pattern of final lengthening, which took into account both final lengthening and shortening of a preceding syllable. As will be explained in section 4.3.3., this grouping was used to examine the production-perception link, specifically focusing on the differences in the onset of final lengthening. For auditory stimuli that include final lengthening in TG1, individuals who produced an early onset of final lengthening (Group 1) are expected to be more sensitive to lengthening, and therefore are predicted to fixate on the word boundary image less and/or more slowly than individuals with a late onset of final lengthening (Group 2). When TG1 has no lengthening, however, individuals in Group 1 are predicted to fixate on the word boundary more and/or more quickly than individuals in Group 2.

4.2. Modelling the production and perception data using GAMMs

For all GAMMs presented in this section, the acoustic values for the three IP boundary properties from the production study are included in the models as smooth terms separately for each type of manipulation applied to the auditory stimuli. The analyses in this chapter used the mgcv package and the visualization of the data used the ggplot2 and the itsadug packages in R.

4.2.1. Pause duration

This analysis examined whether participants' target fixations in response to auditory stimuli with a pause between TG1 and TG2 systematically varied depending on those participants' produced pause durations. A series of three GAMM was fitted, for each of the three manipulation types that included pause duration at the target boundary (conditions M9, M10, M11), with a smooth function which modeled the effect of pause duration on the shape of the proportion fixation over time. A tensor product interaction term was included to specify the interaction between Time and Pause Duration. The structure and output of the models are given in Tables 4.1 to 4.3.

Model Structure:				
Proportion fixation on IP boundary ~ s(Time) + s(Pause Duration) + ti(Time, Pause Duration) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.14257	0.03552	4.014	5.99e-05 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	6.221	6.848	23.817	<2e-16 ***
s(pausedur)	1.000	1.000	0.461	0.497
ti(Time,pausedur)	1.000	1.000	0.349	0.555
s(Time, Participant)	131.098	169.000	30.054	<2e-16 ***

Table 4.1. Structure and output of the Pause Duration model for stimuli with early onset final lengthening (M9).

Model Structure:				
Proportion fixation on IP boundary ~ s(Time) + s(Pause Duration) + ti(Time, Pause Duration) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.15659	0.03218	4.866	1.15e-06 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	6.400	7.004	22.812	<2e-16 ***
s(pausedur)	1.000	1.000	0.356	0.551
ti(Time,pausedur)	4.169	4.373	0.843	0.465
s(Time, Participant)	131.120	169.000	27.648	<2e-16 ***

Table 4.2. Structure and output of the Pause Duration model for stimuli with late onset final lengthening (M10).

Model Structure:				
Proportion fixation on IP boundary ~ s(Time) + s(Pause Duration) + ti(Time, Pause Duration) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.16069	0.03529	4.553	5.33e-06 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	5.937	6.535	20.383	<2e-16 ***
s(pausedur)	1.000	1.000	0.408	0.523
ti(Time,pausedur)	1.001	1.002	0.672	0.412
s(Time, Participant)	134.531	169.000	38.681	<2e-16 ***

Table 4.3. Structure and output of the Pause Duration model for stimuli with late onset final lengthening + S1 shortening (M11).

The summary of the models indicated that the main effect of Pause Duration is not significant (p=.50 for M9; p=.55 for M10; p=.52 for M11), nor is the interaction between Time and Pause Duration significant (p=.56 for M9; p=47 for M10; p=.41 for M11). That is, individual participants' average produced pause durations do not predict the time course of their perceptual fixations in response to auditory stimuli with pause. Figure 4.2 shows the three-dimensional surfaces for each manipulation type. The prediction for the Pause Duration models is that individuals with longer pause durations will respond earlier than those with shorter pause durations. While the effect of individuals' pause durations is not statistically significant, the leftward skew in the three surfaces suggests that there might be a slight trend in which participants who produced longer pause durations (higher end of the y-axis) fixated on the IP boundary image earlier than those who produced shorter pause durations on average (lower end of the y-axis).

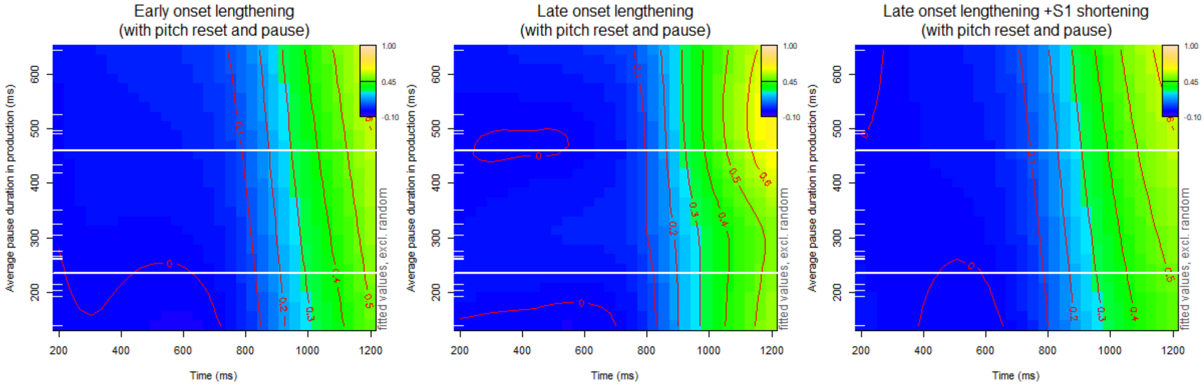


Figure 4.2. Three-dimensional surfaces for the interaction between Pause Duration and Time for the three lengthening conditions. The z-axis represents proportion fixations on the IP boundary image (warmer colors: more target fixations). The white ticks on the y-axis represent individuals' average pause durations in the IP boundary condition. The white horizontal lines indicate the 25th and 75th percentiles of pause durations.

Figure 4.3 shows the modeled proportion fixations on the IP boundary image at the 25th and 75th percentiles of pause durations (see white horizontal lines in Figure 4.2) and the smooths that modeled the difference between them. There is a very weak trend for the estimate at the longer pause duration (75th percentile; green) to show earlier target fixations than the estimate at the shorter pause duration (25th percentile, yellow). The individual smooths shown in Figure 4.4 further illustrate the trend observed in Figures 4.2 and 4.3.

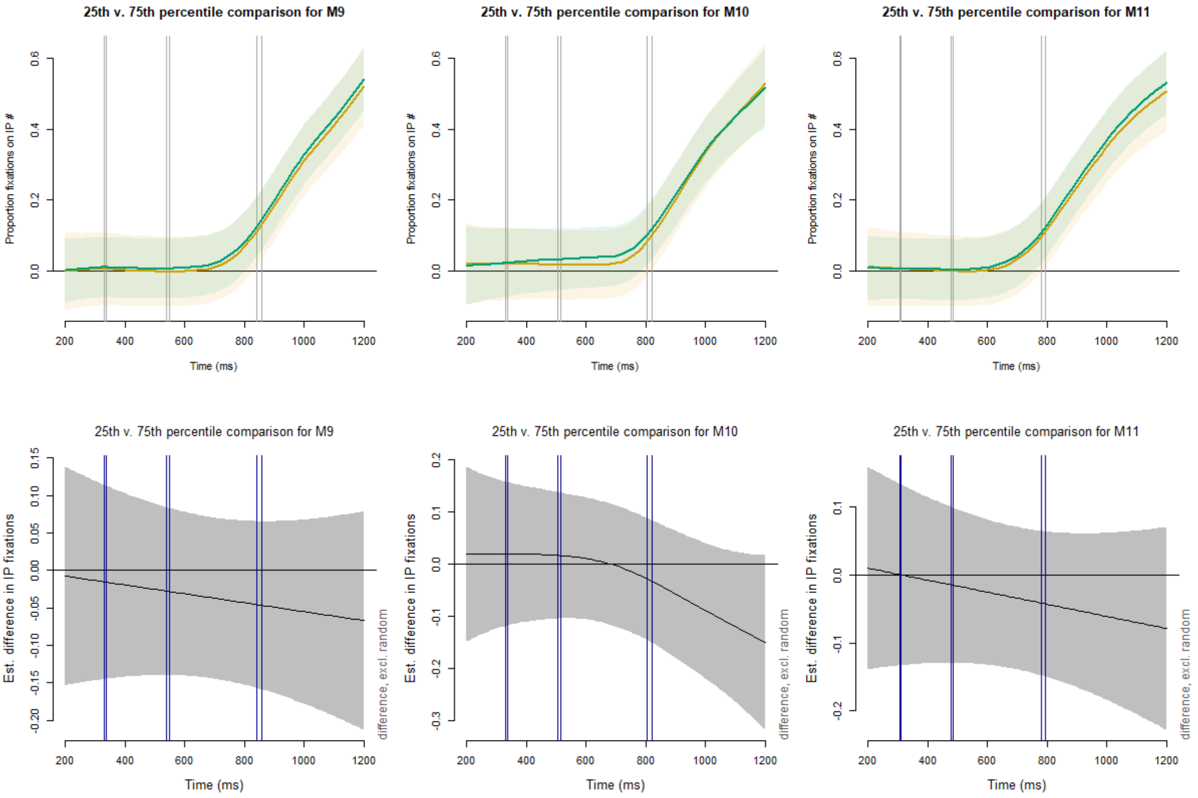


Figure 4.3. Estimated proportion fixations on the IP boundary image over time for three lengthening patterns at the 25th (yellow) and 75th (green) percentiles for pause duration. Bottom plots: difference smooths that modeled the difference between the predicted smooths for the three lengthening patterns.

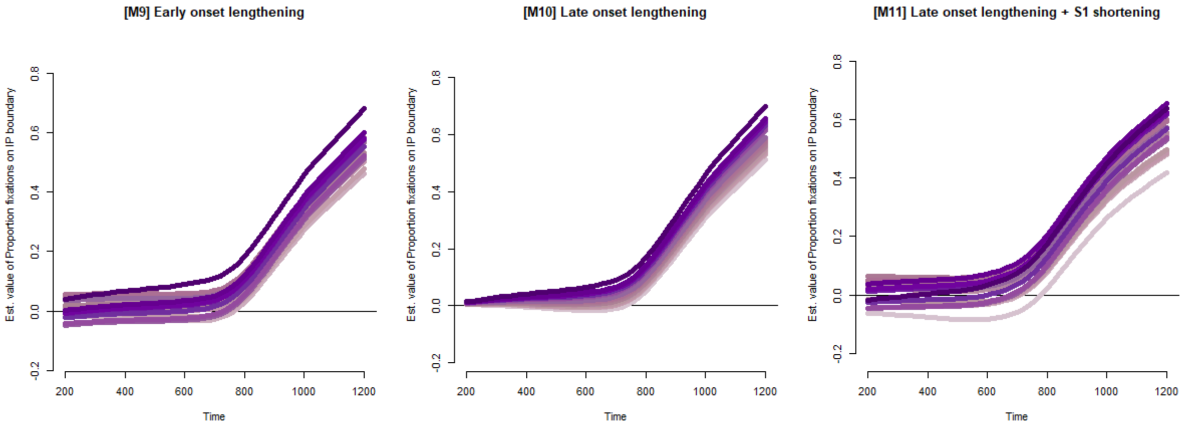


Figure 4.4. Individual proportion fixations on the IP boundary image over time. Each smooth represents the model's predicted perceptual responses of an individual listener. Color of the smooths corresponds to the Y-axis of Figure 4.2 (lighter shade = shorter produced pause duration, darker shade = longer produced pause duration).

4.2.2. Pitch reset

The average f0 maximum difference in the IP boundary condition for each of the 19 participants was included in the GAMM as a smooth function (Pitch Reset) as well as a part of a tensor product interaction with Time. The model structure and output are given in Tables 4.4 through 4.6.

Model Structure:				
Proportion fixation on IP boundary ~ s(Time) + s(f0diff_hz) + ti(Time, f0diff_hz) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.24230	0.02755	8.794	<2e-16 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	6.036	6.679	27.171	<2e-16 ***
s(f0diff_hz)	1.000	1.000	0.085	0.771
ti(Time, f0diff_hz)	4.071	4.329	0.335	0.874
s(Time, Participant)	124.920	169.000	19.462	<2e-16 ***

Table 4.4. Structure and output of the Pitch Reset model for stimuli with no pitch reset (M0).

Model Structure:				
Proportion fixation on IP boundary ~ s(Time) + s(f0diff_hz) + ti(Time, f0diff_hz) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.2447	0.0266	9.2	<2e-16 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	6.311	6.965	28.650	<2e-16 ***
s(f0diff_hz)	1.003	1.003	0.046	0.829
ti(Time, f0diff_hz)	5.934	6.248	1.425	0.166
s(Time, Participant)	122.460	169.000	17.229	<2e-16 ***

Table 4.5. Structure and output of the Pitch Reset model for stimuli with small pitch reset (M1).

Model Structure:				
Proportion fixation on IP boundary ~ s(Time) + s(f0diff_hz) + ti(Time, f0diff_hz) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.23972	0.02348	10.21	<2e-16 ***
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	7.049	7.729	41.880	<2e-16 ***
s(f0diff_hz)	1.706	1.736	0.604	0.472
ti(Time, f0diff_hz)	1.406	1.422	0.207	0.753
s(Time, Participant)	116.503	169.000	14.416	<2e-16 ***

Table 4.6. Structure and output of the Pitch Reset model for stimuli with large pitch reset (M2).

As a reminder, the prediction for the Pitch Reset models is that individuals who produce larger f0 differences will fixate on the word boundary image earlier than those with smaller f0 differences if the stimuli do not have any of the three IP boundary markers. However, if the stimuli contain a small or large pitch reset in the absence of final lengthening or pause, individuals with larger f0 differences may be especially likely to detect a conflict between the presence of pitch reset conflicting with the absence of the other IP boundary markers, and therefore respond more slowly than those with smaller f0 differences. The modeled results show that (participants' produced) Pitch Reset (shown as f0diff_hz) is not predicted to have a significant influence on word boundary fixations (p=.77 for M0, p=.83 for M1, p=.47 for M2). The interaction between Pitch Reset and Time is also not significant (p=.87 for M0, p=.17 for M1, p=.75 for M2), suggesting that the individuals' average pitch reset values do not influence the modeled fixations over time. The contour plots for the interaction shown in Figure 4.5 allow visual inspection of the pattern of the responses. It appears that participants' average pitch resets in production across the IP boundary do not influence in a systematic way how participants fixate on the word boundary image when the stimuli contain no reset (M0; left panel) or a large reset (M2; right panel). When the stimuli contain a small reset (M1; middle panel), it appears that there may be a weak trend for a leftward skew after about 800ms. The modeled predictions for the 25th and 75th percentiles of produced f0 difference and the difference smooths are given in Figure 4.6. Despite the non-significance of the difference, there is a weak trend in which the estimate at the larger f0 difference (75th percentile; green) shows earlier target fixations than the estimate at the smaller f0 difference (25th percentile, yellow) after pitch reset information becomes available (at TG1 offset, indicated by the last pair of vertical lines in the figure). The individual predicted proportion fixations in Figure 4.7 show a similar slight trend for individuals with larger f0 differences compared to those with smaller f0 differences, but in conditions M1 and M2 where a small or large pitch reset was present, this weak trend is not in the predicted direction.

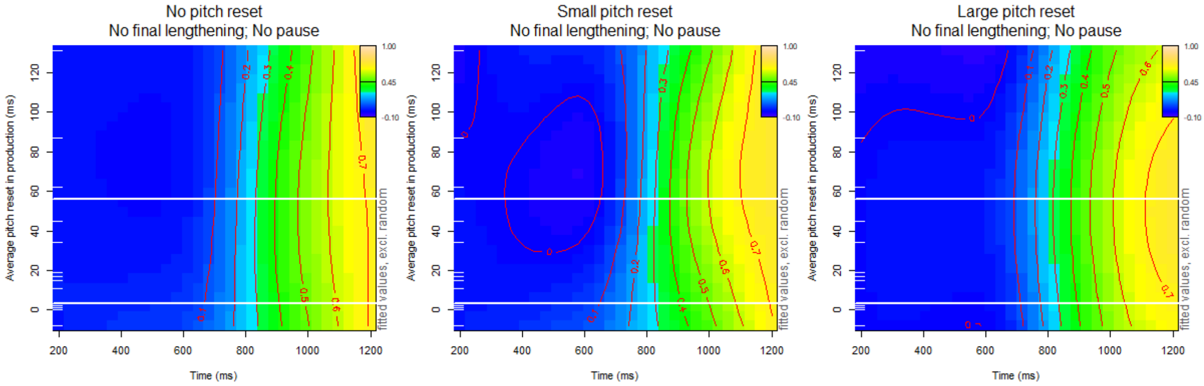


Figure 4.5. Three-dimensional surfaces for the interaction between Pitch Reset and Time for the three pitch reset conditions. The z-axis represents proportion fixations on the word boundary image (warmer colors: more target fixations). The white ticks on the y-axis represent individuals' average f0 difference across the IP boundary. The white horizontal lines indicate the 25th and 75th percentiles of pitch reset values in production.

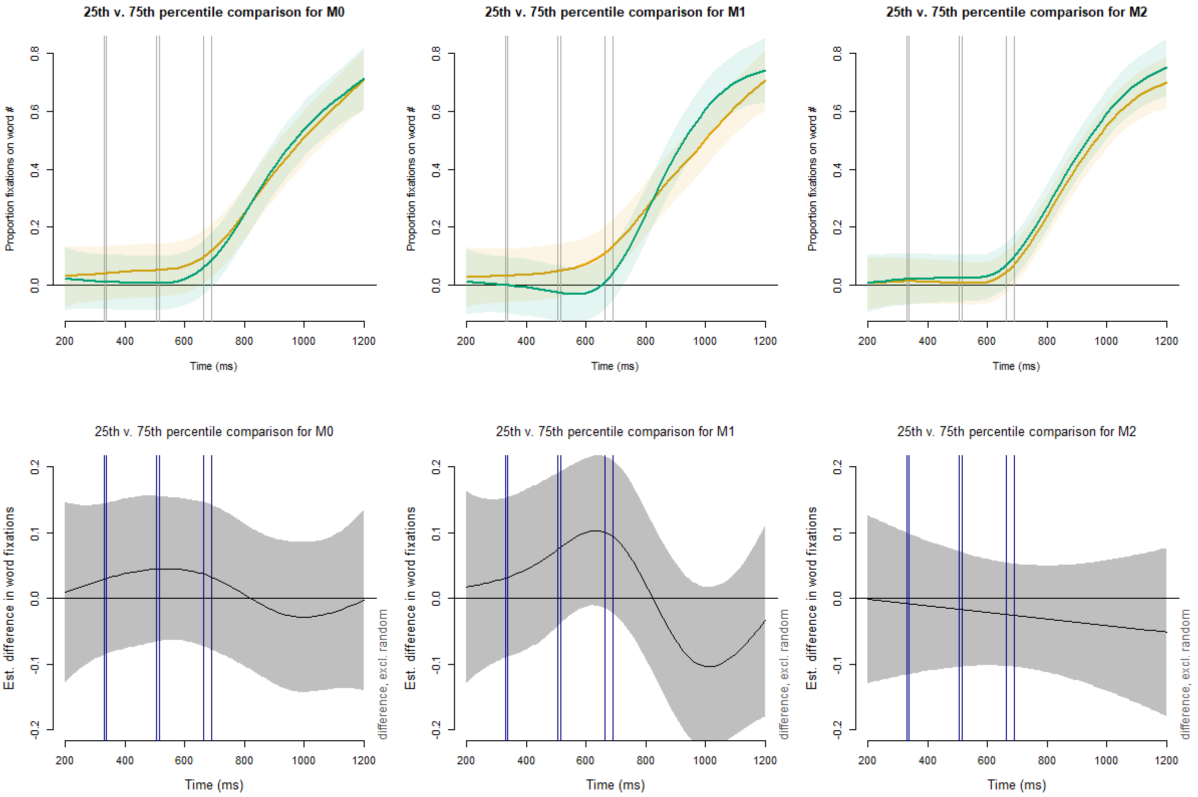


Figure 4.6. Estimated proportion fixations on the word boundary image over time for three pitch reset conditions at the 25th (yellow) and 75th (green) percentiles for f0 difference. Bottom plots: difference smooths that modeled the difference between the predicted smooths for the three pitch reset conditions.

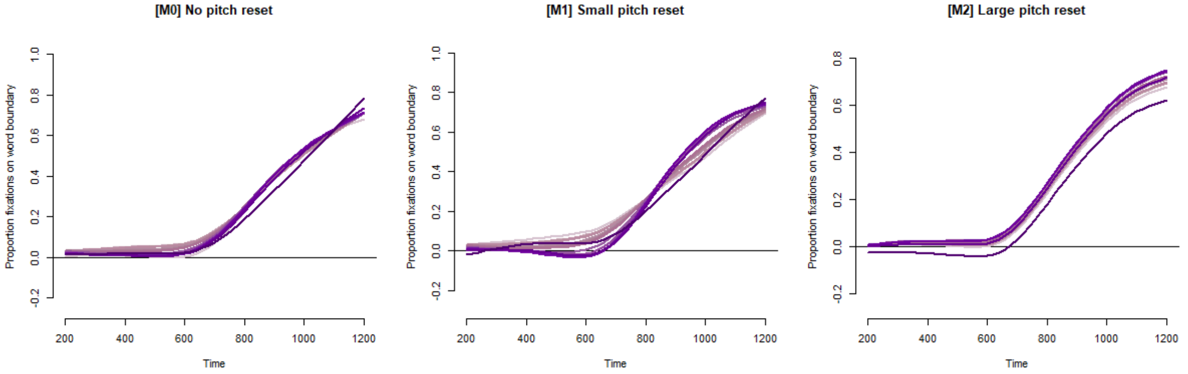


Figure 4.7. Individual proportion fixations on the word boundary image over time. Each smooth represents the model's predicted perceptual responses of an individual listener. Color of the smooths corresponds to the y-axis of Figure 4.5 (lighter shade = smaller f0 difference, darker shade = larger f0 difference).

4.2.3. *Final lengthening*

Unlike the previous models, which incorporated the continuous production values of the IP marker, the Final Lengthening model could not use a continuous production value as the predictor variable, because the measures used to examine the different patterns of produced temporal modulation are for an entire syllable – lengthening of S2 and S3 of TG1 and shortening of S1 – and so are inherently discontinuous. Instead, the Final Lengthening model used the grouping based on the syllable-based patterns of final lengthening established in Chapter 2 (Figure 2.18). Table 4.7. summarizes the produced lengthened patterns of the 19 participants in the perception experiment. Of the 19 participants who returned and whose perceptual data were included in the analyses, nine were from Group 1, six from Group 2, three from Group 3, and one from Group 6. No participant from Group 4 or 5 returned.

Participant number in production study	Grouping	Final lengthening pattern
P31	Group 1	S2 and S3 lengthened
P29	Group 1	
P21	Group 1	
P27	Group 1	
P12	Group 1	
P14	Group 1	
P16	Group 1	
P22	Group 1	
P13	Group 1	
P15	Group 2	S3 lengthened
P10	Group 2	
P09	Group 2	
P28	Group 2	
P06	Group 2	
P04	Group 2	
P08	Group 3	S3 lengthened + S1 shortened
P17	Group 3	
P11	Group 3	
P26	Group 6	S2 and S3 lengthened + S1 shortened

Table 4.7. Participant numbers in production and perception studies and the grouping based on the final lengthening pattern.

Using this grouping does not capture the different absolute durations of final lengthening (i.e., the amount of slowing down at the IP boundary) at the individual level, but rather reflects the different scopes of final lengthening and the presence or absence of preceding shortening observed in the production study. Due to the small sample sizes of Groups 3 and 6, only the data from Groups 1 and 2 were included in the GAMM in which Group was included as a categorical predictor variable that tested whether Group (information from production) predicts participants' perceptual use of temporal information for an upcoming boundary.

The model structure and output are presented in Tables 4.8 through 4.11. As a reminder, it is predicted that individuals with earlier onset of lengthening in TG1 (Group 1) would be more sensitive to the presence of final lengthening than individuals with later onset of lengthening (Group 2). Therefore, compared to the individuals in Group 2, the individuals in Group 1 are expected to fixate less and/or more slowly on the word boundary image when there is lengthening (M3, M5, M7), but fixate more and/or more quickly on the word boundary image when there is no lengthening (control condition; M0).

<table>
<tr><td colspan="5">Model Structure:</td></tr>
<tr><td colspan="5">Proportion fixation on word boundary ~ IsGroup1 + s(Time) + s(Time, by=IsGroup1) + s(Time, Participant)</td></tr>
<tr><td colspan="5">Parametric coefficients:</td></tr>
<tr><td></td><td>Estimate</td><td>Std. Error</td><td>z value</td><td>Pr(>|z|)</td></tr>
<tr><td>(Intercept)</td><td>0.27471</td><td>0.03834</td><td>7.166</td><td>8.19e-13 ***</td></tr>
<tr><td>IsGroup1</td><td>-0.02575</td><td>0.05382</td><td>-0.478</td><td>0.632</td></tr>
<tr><td colspan="5">Approximate significance of smooth terms:</td></tr>
<tr><td></td><td>edf</td><td>Ref.df</td><td>F</td><td>p-value</td></tr>
<tr><td>s(Time)</td><td>6.057</td><td>6.720</td><td>13.597</td><td><2e-16 ***</td></tr>
<tr><td>s(Time): IsGroup1</td><td>1.001</td><td>1.002</td><td>0.009</td><td>0.927</td></tr>
<tr><td>s (Time, subj)</td><td>96.670</td><td>133.000</td><td>16.503</td><td><2e-16 ***</td></tr>
</table>

Table 4.8. Structure and output of the Final Lengthening model with the stimuli containing no final lengthening (M0).

Model Structure:				
Proportion fixation on word boundary ~ IsGroup1 + s(Time) + s(Time, by=IsGroup1) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.123763	0.041908	2.953	0.00315 **
IsGroup1	-0.002849	0.056673	-0.050	0.95991
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	5.750	6.388	19.309	<2e-16 ***
s(Time): IsGroup1	1.001	1.001	0.002	0.962
s (Time, subj)	98.565	133.000	17.116	<2e-16 ***

Table 4.9. Structure and output of the Final Lengthening model with the stimuli containing the early onset of final lengthening (M3).

Model Structure:				
Proportion fixation on word boundary ~ IsGroup1 + s(Time) + s(Time, by=IsGroup1) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.15275	0.05338	2.862	0.00422 **
IsGroup1	0.01595	0.07135	0.224	0.82314
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	5.933	6.582	18.129	<2e-16 ***
s(Time): IsGroup1	1.002	1.002	0.002	0.968
s (Time, subj)	98.646	133.000	20.695	<2e-16 ***

Table 4.10. Structure and output of the Final Lengthening model with the stimuli containing the late onset of final lengthening (M5).

Model Structure:				
Proportion fixation on word boundary ~ IsGroup1 + s(Time) + s(Time, by=IsGroup1) + s(Time, Participant)				
Parametric coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.13161	0.05021	2.621	0.00878 **
IsGroup1	0.07266	0.06695	1.085	0.27779
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Time)	6.372	7.074	19.630	<2e-16 ***
s(Time): IsGroup1	1.001	1.002	1.121	0.29
s (Time, subj)	93.708	133.000	16.081	<2e-16 ***

Table 4.11. Structure and output of the Final Lengthening model with the stimuli containing the late onset of final lengthening + S1 shortening (M7).

The results indicate that the effect of Group is not significant in any of the four conditions (p=.63 for M0; p=.96 for M3; p=.82. for M5; p=.28 for M7). The effect of Group on proportion fixation over time is also not significant (p=.93 for M0; p=.96 for M3; p=.97 for M5; p=.29 for M7). Figure 4.8 shows the model predictions for Group 1 and 2 in the four lengthening conditions, and Figure 4.9 presents the difference smooths that modeled the difference between the estimates for Group 1 and Group 2 in the four conditions.

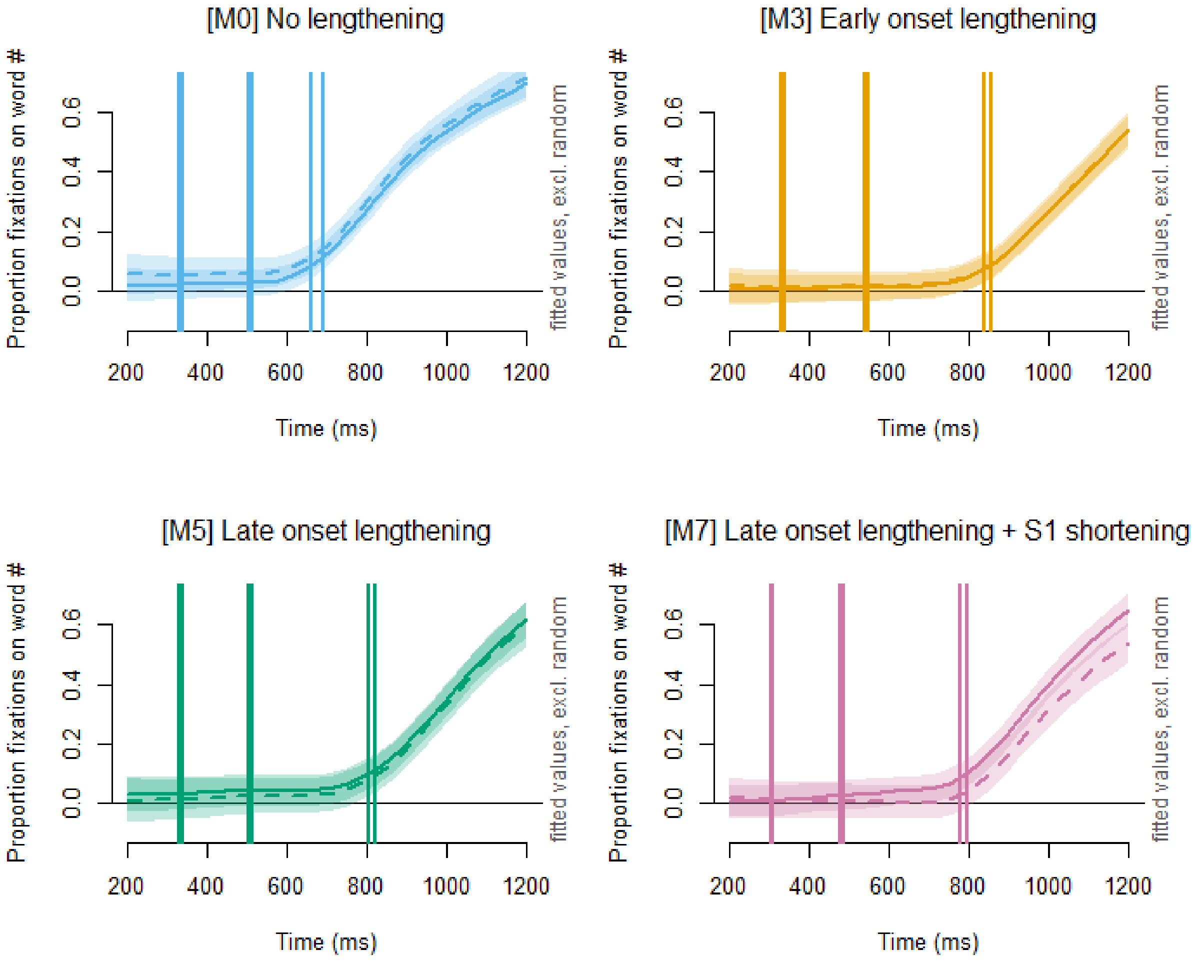


Figure 4.8. Model predictions for the three lengthening patterns and the control condition in the absence of pitch reset and pause between TG1 and TG2. The x-axis represents the time range for the data used in the model (perceptual responses near 100% after 1200ms). The pairs vertical lines are syllable boundaries of TG1. Solid lines = Group 1 (early onset of lengthening); Dashed lines = Group 2 (late onset of lengthening).

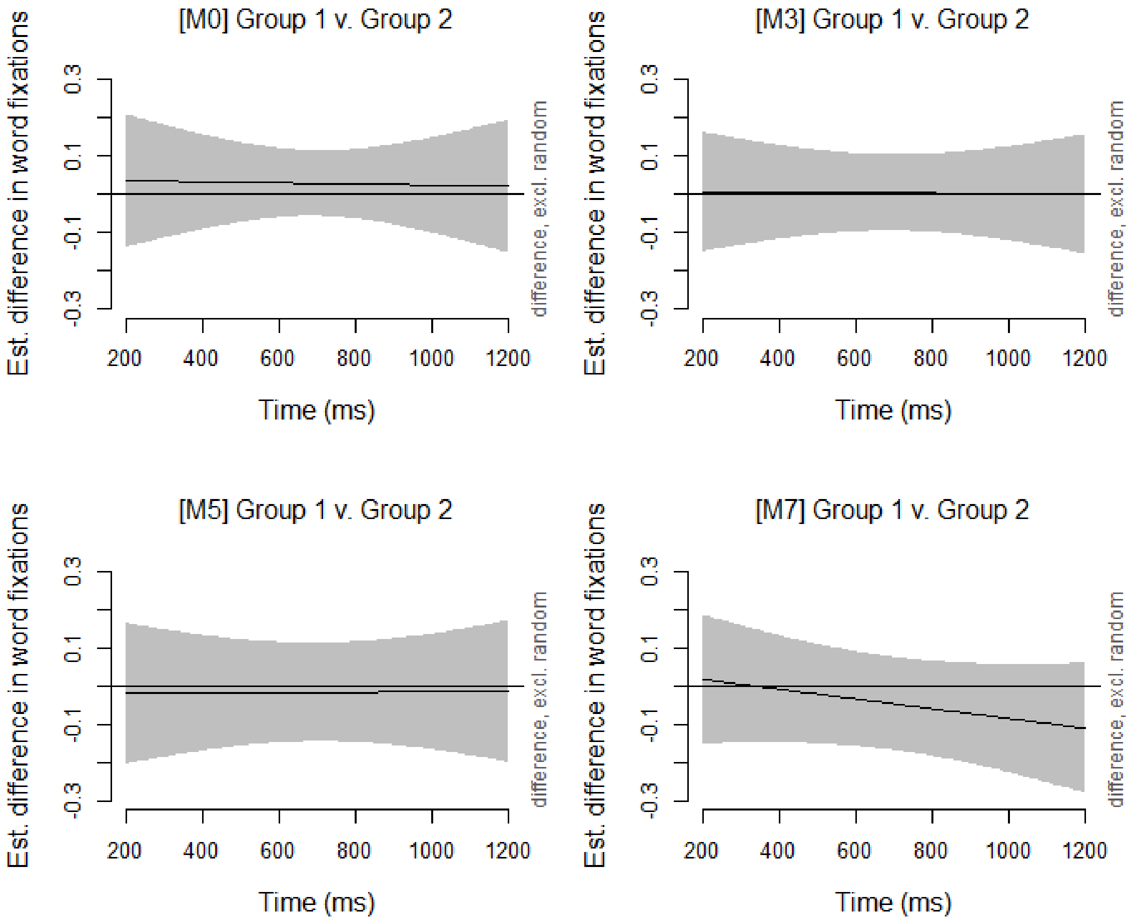


Figure 4.9. Difference smooths for the final lengthening patterns (M3, M5, M7) and the control condition (M0).

Contrary to the prediction, individuals in Group 1 and Group 2 did not differ in their perceptual responses to the stimuli with or without final lengthening. There was a weak trend observed after the S2 boundary in the condition with late onset of lengthening and S1 shortening (M7) in which Group 1 tended to fixate on the word boundary image earlier and (overall) more than Group 2.

4.3. Discussion

In each subsection, the results of the GAMM analyses are discussed in terms of the predictions for the relation between individuals' production of the acoustic properties of IP boundary and individuals' perception of those properties.

4.3.1. Pause duration

Pause durations measured in the IP boundary condition in the production experiment showed inter-speaker variation: the 19 participants whose perception data were also analyzed produced average pause durations that were as short as 139ms and as long as 643ms. Under the general hypothesis that perceptual use of information for prosodic boundaries reflects production of that information, it was predicted that, when hearing stimuli with a pause between TG1 and TG2, individuals who produce longer pause durations will be faster to identify the target boundary as an IP boundary than individuals with shorter pause durations.

The model did not find that participants' produced pause durations were a significant predictor for their proportion fixations on the IP boundary image over time. However, a similar leftward skew was observed in all three lengthening conditions (Figure 4.2), and the direction of the trend was consistent with the prediction. Figure 4.4 showed that, while there is considerable overlap between individual participants' modeled proportion fixations over time, individuals with longer pause durations (darker shade) tended to fixate the IP boundary image earlier than those with shorter pause durations (lighter shade).

It is worth noting that participants heard only one duration of pause in the experiment. Given the variation of pause duration observed in the production study, individuals' perceptual biases may manifest to a greater extent if participants were given varying durations of pause in the auditory stimuli.

4.3.2. Pitch reset

It was predicted that, in the absence of any IP boundary marker – including pitch reset – in the auditory stimuli, participants who produce larger f0 differences in the IP boundary condition would fixate earlier on the word boundary image than those who produce smaller f0 differences. On the other hand, participants with larger f0 differences were expected to find the presence of a small or large pitch reset in the absence of two other IP markers confusing, thus responding more slowly than those with smaller f0 differences who are expected to be less sensitive to the intonational cue.

The results of the Pitch Reset model do not align with the prediction, in that the effect of Pitch Reset was not significant. Although there was a very weak trend in all three reset conditions for participants with larger pitch reset values in production to fixate the target earlier than those with smaller values after reset information became available (at TG1 offset), the conditions with slightly less overlap in the individual smooths (Figure 4.7) were not in the predicted direction. There remains an empirical question as to whether a different measure for pitch reset, such as difference in pitch range across target boundary, would provide a clearer pattern in either direction.

4.3.3. Final lengthening

The predictions for the four conditions with three different lengthening patterns and one control (no lengthening) stated that individuals in Groups 1 and 2 will respond differently in all four conditions, based on the assumption that individuals who produced early onset of lengthening (i.e., larger scope of lengthening) would be more sensitive to the presence of final lengthening than those who produced late onset of lengthening. However, the results did not support the predictions, in that individuals' perceptual responses did not differ depending on the scope of the final lengthening they produced.

For the control condition, with no final lengthening or other acoustic information for an IP boundary (M0), failure to support predictions may be due to absence of any conflicting information that might have triggered production-related cue weightings. However, when final lengthening was present in TG1 in the absence of pitch reset and pause – i.e., when listeners received conflicting information about the upcoming boundary – the lack of difference between the groups would seem to suggest that the conflict, at least as captured in these stimuli, was not systematically more or less disruptive depending on these listeners' own lengthening patterns.

Additional analysis is needed to determine whether the effects and tendencies may be present when both production and perception data are analyzed in a more fine-grained manner. For example, individual participants' production and perception of final lengthening may be analyzed segment-by-segment rather than syllable-by-syllable which may provide a more detailed look into the individual-specific pattern of producing and perceiving final lengthening.

4.4. Concluding remarks and limitations

A main goal of the study is to investigate the role of individual differences in the relation between production and perception of prosodic boundaries. The results are not conclusive, with lack of statistical significance for many of the effects of the production information as predictor variables. This may be due at least in part to the relatively small sample size: only 20 out of 32 participants returned for the perception study and one participant's perception data were not used for analysis because their eye-movements could not be reliably tracked. Furthermore, the manipulated pitch reset might not have been salient enough for the majority of listeners, given that about half of the speakers' average pitch reset was larger than the model speaker’s reset values (and hence the value used in the manipulations). The strongest trend observed in the modeled visualizations of the data was for pause duration, which was consistent with predictions for the production-perception relation in all relevant conditions. Chapter 5 discusses the extent to which the study’s findings provide insights into the general hypothesis that speaker-specific production of prosodic boundaries is reflected in the perception of those boundaries.

CHAPTER 5

General discussion and conclusion

This dissertation investigated how speakers produce and listeners perceive information for prosodic boundaries in English and whether individual differences in the production of this information are related to differences in perception. An acoustic production experiment and a visual world perception experiment were conducted. The goal of the production study in Chapter 2 was to characterize how individual speakers differ in their production of three primary acoustic properties associated with IP (as compared to word) boundaries: pause, pitch reset, and the temporal modulation of the pre-boundary (TG1) and post-boundary (TG2) word. The goal of the perception study in Chapter 3 was to test how listeners use these properties in real time, as they unfold in the acoustic signal, to differentiate IP and word boundaries. In Chapter 4, the production results were used to model the perception data to determine whether and how speaker-specific patterns of producing the properties might be reflected in the same individuals' perception of IP boundaries.

5.1. Production overview

The hypotheses that motivated the production study were that all speakers would produce one or more of the three type of acoustic boundary information, but that individual speakers will show substantial variation as to the combination of the properties used to distinguish IP and word boundaries as well as the degree to which they were used. The results of the production study found that, as a group, speakers marked an IP boundary with pitch reset, pause, lengthening of two syllables preceding the pause, and shortening of the first syllable of TG2. Individually, the 32 participants consistently employed some combination of pause, pitch reset, and final lengthening to distinguish IP boundaries from word boundaries.

The results also demonstrated that, in line with the prediction, individual speakers varied to a substantial degree in how they acoustically realized the IP boundary information. For example, their average pause durations and f0 maximum differences measured at IP boundaries exhibited a gradient pattern. The average pause durations ranged from 139ms to 643ms, and the average f0 differences for the 20 participants who used this information to differentiate IP from word boundaries ranged from 1Hz to 131Hz, while 11 participants did not produce an f0 difference between TG1 and TG2 to signal IP boundaries, and one participant produced a negative difference at IP boundaries instead of a positive reset. The final lengthening analysis classified the 32 participants into six different groups depending on the syllabic scope of final lengthening and the presence of pre-boundary shortening, characterizing the speaker-specific temporal modulation associated with IP boundaries. The majority of speakers showed a leftward spreading of the boundary-related lengthening effect (i.e., lengthening of the last or last two syllables of the pre-boundary word), reinforcing the findings of previous studies that reported similar results (Berkovits, 1994; Cambier-Langeveld, 2000; Fougeron & Keating, 1997; Shattuck-Hufnagel & Turk, 1998; Byrd et al., 2006; Turk & Shattuck-Hufnagel, 2007; Katsika, 2016). Some speakers exhibited shortening of a syllable of TG1 (S1 or S2) and/or TG2 (S4). The boundary-related temporal modulations are consistent with the predictions of the π-gesture model (Byrd & Saltzman, 2003), namely that the effect of the boundary extends over a period of time and is local and continuous. The source of the individual differences in temporal effects may be the individually variable onset and scope of the π-gesture, and/or in the variability of coupling strength between the π-gesture and the μ-gesture (as suggested in Katsika et al., 2014). A more fine-grained analysis and computational modeling of the speech data are needed to test these possible explanations.

Tests for correlations between the type and extent of boundary information produced by individual speakers found significant trading and enhancing relations between pause duration, f0 difference, and phrase-final syllable (S3) duration. For example, a trading relationship was found between pause duration and S3 duration for five participants, in line with the findings in Ferreira, 1993. Another five participants showed a trading relation between f0 difference and S3 duration. An enhancement relationship was found between pause duration and pitch reset for six participants, and one participant showed this relationship between pause duration and S3 duration. Overall, though, there was no single or even predominant way that emerged by which speakers used one IP marker in relation to another.

5.2. Perception overview

The eye-tracking experiment first investigated whether listeners, as a group, are sensitive to different combinations and degrees of these IP markers when they are independently manipulated in the auditory stimuli. Consistent with predictions, listeners found pause duration to be a salient cue for IP boundary, showing a robust difference in their final fixations depending on the presence or absence of pause duration: listeners overwhelmingly looked to the IP boundary image when a pause was present between TG1 and TG2, and to the word boundary image when no pause was present. This finding is in line with previous studies that demonstrated the role of pause as a salient cue for an IP boundary (Swerts & Geluykens, 1994; de Pijper & Sanderman, 1994; Zhang, 2012; Roy et al., 2017, Petrone et al., 2017). Final lengthening was also used by the listeners to identify an upcoming prosodic boundary: listeners fixated earlier and more (in the earlier time course of the trial) on the word boundary image when they heard stimuli with no lengthening than when they heard stimuli with lengthening in TG1. However, presence of final lengthening did not lead listeners to fixate on the IP (rather than word) boundary image, with or without pitch reset, suggesting that the effect of lengthening – at least of the magnitude used in this study – was to delay listeners' boundary decisions rather than to shift their percepts from one boundary type to the other. This relatively small effect of final lengthening does not corroborate studies that suggested a more heavily weighted role of final lengthening (e.g., Lehiste et al., 1976; Scott, 1982; Petrone et al., 2017), and is a surprising result considering how prevalent lengthening is in American English. It might be that the bias for the word boundary image in participants' perceptual responses is due to the ambiguous role of lengthened syllable durations, which can cue both prominence as well as boundary (although the contrastive focus was placed on a word that occurs earlier in the sentence and not on TG1). Further investigation of perceptual responses for the same stimuli when TG1 received contrastive focus will shed light on the presence and source of the bias.

Unlike presence of pause or final lengthening in the auditory stimuli, the pitch reset manipulation did not yield a clear perceptual pattern. In the absence of pause and final lengthening, contrary to expectations, listeners did not fixate earlier and more on the word boundary image when the stimuli had no pitch reset than when stimuli had a small or large reset. Listeners also did not generally fixate more on the word boundary image when the stimuli had small rather than large reset. However, the results for the manipulations with pitch reset, lengthening, and no pause (section 3.3.2.2) showed an effect of pitch reset in the early onset lengthening condition (Figure 3.10), where listeners were more likely to fixate on the word boundary image when the stimuli had no pitch reset compared to when it had small reset, suggesting that the absence of pitch reset facilitates the percept of a word boundary in the presence of what seems to be strongly conflicting information. The small and inconsistent effect of the f0 cue is also not consistent with its larger role found in some studies (Ladd, 1988; Seidl, 2007; Bögels & Torreira, 2015), and is more consistent with studies finding a relatively minor role of pitch in boundary perception (e.g., Swerts, 1997; Petrone et al., 2017; Roy et al., 2017).

One major finding is that listeners did not wait until all boundary information became available to look at the target image, but rather began to fixate on the target nearly as soon as reliable information about the target boundary became available. The effect of conflicting information about the type of boundary induced differences in fixations at the relevant time ranges (such as the one described above, for example), although the differences were not always consistent with the direction of the predictions. The results of the group analysis contribute to the literature on weighting of the primary acoustic properties for an IP boundary by characterizing the time course of the perception of different combinations of these properties (cf. Lee et al., 2008).

5.3. The production-perception relation

A main goal of the eye-tracking study was to investigate whether the individual participants' production patterns are reflected in their perception. Under the general hypothesis that production and perception of prosodic boundaries are closely related within individuals, it was predicted that participants' perceptual responses to auditory stimuli containing different combinations of the IP boundary properties would reflect their own production of these properties. The output of a series of perceptual models that incorporated information about individuals' production of the relevant properties did not support the hypothesis. The weak trends observed in the pitch reset model were mostly not consistent with predictions, and the final lengthening model did not provide any trend that was consistent with predictions. The strongest support for the hypothesis concerning the production-perception relation was that the trend for the pause duration model, across the three lengthening conditions, were all consistent with the prediction that, when pause is present, individuals who produce longer pauses are faster to fixate on the IP boundary image than individuals who produce shorter pauses.

Implications can be drawn from the production and perception studies. Pitch reset as a marker for an IP boundary was less consistently produced and less perceptually useful than pause and final lengthening. For production, not all participants produced pitch reset, whereas all participants produced pause and final lengthening. For perception, pitch reset did not induce perception of an IP boundary in a way that is as robust as pause or final lengthening. This is consistent with previous studies that showed substantial individual variation in phrase-final f0 events (Swerts, 1997; Zhang, 2012; Petrone et al., 2017).

In addition, despite the lack of clear evidence of a close relation between production and perception of prosodic boundaries, both the production and perception studies demonstrate individual differences to substantial degrees, adding to the body of research showing individual differences in production and perception of prosodic boundaries or prominences, which suggest that the mixed findings on the weighting of the cues for prosodic boundaries discussed in section 5.2 may be explained in part by individual differences.

The lack of clear evidence for a close relation between production and perception is not entirely surprising. Many of the previous studies that have investigated the relation between production and perception of segmental variation found no or only weak supporting evidence (e.g., Grosvald, 2009 and Grosvald & Corina, 2012 for vowel-to-vowel coarticulation; Shultz et al., 2012, Schertz et al., 2015, and Coetzee et al., 2018 for the relation between f0 and Voice Onset Time). This relation has not been previously examined for production and perception of prosodic boundaries within the same individuals.

Chapter 1 motivated this investigation in part by an interest in the phonetic sources of sound change and especially in the assumption of accounts of perceptually motivated changes that listeners' perceptual strategies are reflected in their own productions (e.g., Ohala 1981). The at best weak relation between individuals' production and perception of prosodic boundaries in the current study would seem to call this assumption into question. However, relatively recent discussions of this issue have emphasized that initiation of sound change due to productions that reflect individuals' perceptual bias does not require that all members of the relevant speech community exhibit a close relation between production and perception (Grosvald & Corina, 2012; Stevens & Harrington, 2014; Beddor et al., 2018). For instance, in their study of production and perception of coarticulatory nasalization, Beddor et al. (2018) found that, in general, participants who perceptually attended more to coarticulatory nasalization in the auditory stimuli tended to produce heavier coarticulatory nasalization. However, this relation between production and perception was not observed for all participants in the study; some participants who produced early onset of coarticulatory nasalization were not the listeners who found vowel nasality especially informative in perception. They suggested that those individuals who reliably manifest their perceptual bias in their production might be sufficient to contribute to the initiation of a sound change. It might be speculated that, despite failure to find supporting evidence for a link between production and perception of pitch reset and final lengthening in the current study, some individuals may nonetheless manifest a close production-perception relation. For example, the unmodeled fixation data from Participant 16 (P16), who produced early onset of lengthening, are suggestive of an interesting pattern for production and perception of the three properties for IP boundary.

Figure 5.1 shows this participant's proportion fixations on the word (yellow) and IP (green) boundary images for the auditory stimuli in conditions M3, M5, M7 (i.e., three different lengthening patterns without pause or pitch reset between TG1 and TG2). It appears that P16 fixated on the word boundary image more when TG1 had late onset of lengthening (M5) and late onset of lengthening + S1 shortening (M7) than early onset of lengthening (M3). This suggests that P16 was more sensitive to a conflicting cue in M3, than in M5 and M7. Given that P16 produced final lengthening in S2 and S3, the fewer fixations on the word boundary image when the onset of final lengthening was at S2 (condition M3) is consistent with P16's perceptual use of the temporal information to distinguish the type of the upcoming boundary being influenced by the participant's own production of final lengthening.

Figure 5.2 shows P16's proportion fixations on the IP (green) and word (yellow) boundary images for the auditory stimuli in conditions M9, M10, and M11. These conditions are comparable to M3, M5, and M7, respectively, except that they include pause and small pitch reset. For these conditions, which elicit fixations on the IP boundary image, final lengthening does not present a conflict, and, correspondingly, P16 appears to fixate roughly equally often on the target image in all conditions.

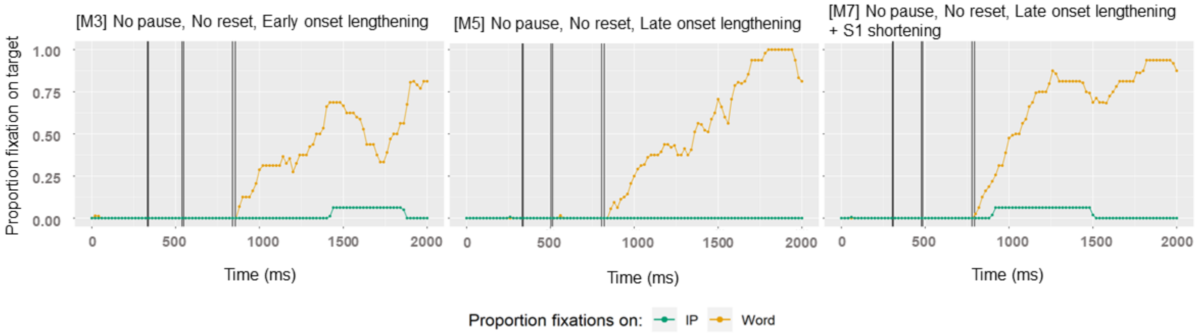


Figure 5.1. Proportion fixations of Participant 16 on the word (yellow) and IP (green) boundary images for stimuli including no pitch reset or pause between TG1 and TG2, for conditions M3 (early onset lengthening), M5 (late onset lengthening), and M7 (late onset lengthening + S1 shortening).

Figure 5.2. Proportion fixations of Participant 16 on the word (yellow) and IP (green) boundary images for stimuli including a small pitch reset and a pause between TG1 and TG2, for conditions M9 (early onset lengthening), M10 (late onset lengthening), and M11 (late onset lengthening + S1 shortening).

This preliminary look at P16's results suggests that, as is the case for some segmental properties, the production-perception link within individuals may also exist for prosodic structure. Further examination of individual participants might shed light on the discussion of how the production-perception relation within individuals could contribute to the initiation of a perceptually-motivated sound change.

It is possible that, at least for certain properties, the current design failed to establish a clear production-perception relation because the acoustic cues used in the auditory stimuli were not sufficiently salient for some participants. For instance, the majority of speakers consistently produced a pitch reset to distinguish IP and word boundaries, and most of these speakers produced a distinction that was larger than the pitch reset produced by the model speaker and subsequently used in the perception experiment. Perhaps, then, the f0 difference used in the auditory stimuli – while carefully done to reflect a large and small reset for the model speaker – needed to be larger to capture its perceptual relevance for the majority of participants. Alternatively, the lack of a robust effect of the presence/absence and size of pitch reset on listeners' perception of prosodic boundaries may suggest that the intonational cue is weighted less heavily than the temporal cues in the perception of an IP boundary.

Further analyses of the existing and new data may shed more light on these and other possible explanations, as well as on the research questions that were met with inconclusive results. For instance, additional analyses could examine the relative weighting of the three properties for the IP boundary within individual participants, to test whether some participants find certain information they consistently produced to differentiate IP and word boundaries more useful than other information they also used. Additional data and analyses could also make use of more fine-grained measures to examine the acoustic properties, a new set of auditory stimuli with more robust manipulation, and more participants.

APPENDIX A

Stimuli for the Production Experiment

(1)	TG1 = ‘naNIna’	# = IP boundary
C: The agent called naNIna. # Navarro and Parker bought the painting.		
T: No, the **painter** called naNIna. # Navarro and **Damon** bought the painting.		
(2)	TG1 = ‘naNIna’	# = word boundary
C: The agent called naNIna # Navarro. And Parker bought the painting.		
T: No, the **painter** called naNIna # Navarro. And **Damon** bought the painting.		
(3)	TG1 = ‘maMIma’	# = IP boundary
C: The paramedic called maMIma. # Melinda and Peter said no one got hurt.		
T: No, the **police** called maMIma. # Melinda and **Danny** said no one got hurt.		
(4)	TG1 = ‘maMIma’	# = word boundary
C: The paramedic called maMIma # Melinda. And Peter said no one got hurt.		
T: No, the **police** called maMIma # Melinda. And **Peter** said no one got hurt.		

(5)	TG1 = ‘naNIna’	# = IP boundary
C: The rancher called naNIna. # Delilah and Paige asked about the apples.		
T: No, the **farmer** called naNIna. # Delilah and **David** asked about the apples.		
(6)	TG1 = ‘naNIna’	# = word boundary
C: The rancher called naNIna # Delilah. And Paige asked about the apples.		
T: No, the **farmer** called naNIna # Delilah. And **David** asked about the apples.		
(7)	TG1 = ‘maMIma’	# = IP boundary
C: The king called maMIma. # Belinda and Paul thought that was rude.		
T: No, the **queen** called maMIma. # Belinda and **Daisy** thought that was rude.		
(8)	TG1 = ‘maMIma’	# = word boundary
C: The king called maMIma # Belinda. And Paul thought that was rude.		
T: No, the **queen** called maMIma # Belinda. And **Daisy** thought that was rude.		

Table A1. Stimuli for the production experiment.

APPENDIX B

Participant Biographical and Language Background Questionnaire

1) Age: _____________
2) Eyesight: Do you wear glasses? (Y / N)
   Do you wear contact lenses? (Y / N)
3) Metal objects:
   Do you wear braces? (Y / N)
   Do you wear fixed retainers? (Y / N)
   Do you wear piercing(s)? (Y / N)
4) Latex allergy:
   Are you allergic to latex? (Y / N)

Our experience with different languages, and with different dialects of English, can influence the way we talk and the way we perceive speech. Having background information on these experiences can be helpful to researchers who study language, although you may choose not to respond to these questions.

1) What language(s) were spoken in your home as a child?
2) What language(s) do you currently speak at home?
3) List other languages with which you have experience, and explain what your experience is. (*For example*: Spanish – 3 years of high school study)
4) List all the places where you have lived for more than 6 months, and indicate the approximate dates that you lived in each of these places.
5) Are you musically trained? Or is a main hobby of yours related to music?
6) Would you be willing to participate in two more experiments for additional compensation?

   (Y / N)

   If yes, we might contact you via email to schedule the following experiments. Because this is a preliminary question, you can always change your mind.

APPENDIX C

Outputs of Statistical Models for Production Experiment

1. Group results (across all 32 participants):

    1.1. Syllable durations

Model: **S1 duration**					
lmer(formula = syl1 ~ boundary + (1 | subj), data = dt)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	312.7	17.68		
Residual		352.6	18.78		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	128.3191	3.1774	31.9981	40.385	<2e-16 ***
boundarywd	1.0427	0.8034	2155.0974	1.298	0.194

Table C1. Structure and output of the model testing the effect of boundary on S1 duration.

Model: **S2 duration**					
lmer(formula = syl2 ~ boundary + target + (1 | subj), data = dt)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	195.4	13.98		
Residual		271.7	16.48		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	169.8997	2.5459	33.5684	66.735	< 2e-16 ***
boundarywd	-7.7480	0.7053	2154.1718	-10.985	< 2e-16 ***
targetnanina	-3.1118	0.7051	2154.0801	-4.414	1.07e-05 ***

Table C2. Structure and output of the model testing the effect of boundary on S2 duration.

Model: **S3 duration**					
lmer(formula = syl3 ~ boundary + target + (1 | subj), data = dt)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	275.0	16.58		
Residual		743.5	27.27		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	243.246	3.101	35.966	78.43	<2e-16 ***
boundarywd	-93.783	1.167	2154.306	-80.39	<2e-16 ***
targetnanina	-15.523	1.166	2154.133	-13.31	<2e-16 ***

Table C3. Structure and output of the model testing the effect of boundary on S3 duration.

Model: **S4 duration**					
lmer(formula = syl4 ~ boundary + target + (1 | subj), data = dt)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	99.23	9.962		
Residual		821.57	28.663		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	74.223	2.058	46.176	36.073	< 2e-16 ***
boundarywd	19.238	1.226	2154.933	15.687	< 2e-16 ***
targetnanina	6.386	1.226	2154.465	5.209	2.08e-07 ***

Table C4. Structure and output of the model testing the effect of boundary on S4 duration.

1.1.1. Significant interactions involving syllable duration

Model: **S2 duration when TG1 is ‘maMIma’**					
lmer(syl2~boundary + (1|subj), data=dtm)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	216.4	14.71		
Residual		246.9	15.71		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	168.4298	2.6864	33.0857	62.696	< 2e-16 ***
boundarywd	-4.8753	0.9544	1053.1742	-5.108	3.86e-07 ***

Table C5. Structure and output of the model testing the effect of boundary on S2 duration when TG1 type is ‘maMIma’.

Model: **S2 duration when TG1 is ‘naNIna’**					
lmer(syl2~boundary + (1|subj), data=dtn)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	203.7	14.27		
Residual		262.6	16.20		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	168.2068	2.6163	33.3743	64.29	<2e-16 ***
boundarywd	-10.6405	0.9775	1069.3063	-10.89	<2e-16 ***

Table C6. Structure and output of the model testing the effect of boundary on S2 duration when TG1 type is ‘naNIna’.

Model: **S2 duration when BOUNDARY is IP boundary**					
lmer(syl2~ target + (1|subj), data=dtip)					
Random effects:					
	Groups	Name	Variance	Std.Dev.	
	subj	(Intercept)	195.9	14.00	
	Residual		208.9	14.45	
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	168.3314	2.5508	32.9366	65.993	<2e-16 ***
targetnanina	-0.2241	0.8741	1062.1031	-0.256	0.798

Table C7. Structure and output of the model testing the effect of TG1 type on S2 duration when boundary is IP boundary.

Model: **S2 duration when BOUNDARY is word boundary**					
lmer(syl2 ~ target + (1|subj), data=dtwd)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	250.5	15.83		
Residual		280.8	16.76		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	163.541	2.889	33.072	56.598	< 2e-16 ***
targetnanina	-5.995	1.015	1060.126	-5.909	4.64e-09 ***

Table C8. Structure and output of the model testing the effect of TG1 type on S2 duration when boundary is word boundary.

Model: **S4 duration when TG1 is ‘maMIma’**					
lmer(syl ~ boundary + (1|subj), data=dtm)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	147	12.12		
Residual		1001	31.64		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	72.659	2.537	42.606	28.64	<2e-16 ***
boundarywd	22.301	1.922	1054.011	11.61	<2e-16 ***

Table C9. Structure and output of the model testing the effect of boundary on S4 duration when TG1 type is ‘maMIma’.

Model: **S4 duration when TG1 is ‘naNIna’**					
lmer(syl4 ~ boundary + (1|subj), data=dtn)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	67.68	8.227		
Residual		626.31	25.026		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	82.118	1.805	45.857	45.48	<2e-16 ***
boundarywd	16.320	1.509	1070.500	10.81	<2e-16 ***

Table C10. Structure and output of the model testing the effect of boundary on S4 duration when TG1 type is ‘naNIna’.

Model: **S4 duration when BOUNDARY is IP boundary**					
lmer(syl4~ target + (1|subj), data=dtip)					
Random effects:					
Groups	Name	Variance	Std.Dev.		
subj	(Intercept)	185.9	13.63		
Residual		504.5	22.46		
Fixed effects:					
	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	72.681	2.596	35.930	27.998	< 2e-16 ***
targetnanina	9.323	1.358	1062.355	6.863	1.14e-11 ***

Table C11. Structure and output of the model testing the effect of TG1 type on S4 duration when boundary is IP boundary.

```
Model: S4 duration when BOUNDARY is word boundary
lmer(syl4 ~ target + (1|subj), data=dtwd)
Random effects:
Groups          Name            Variance     Std.Dev.
subj           (Intercept)        69.63        8.345
Residual                        1079.75       32.860
Fixed effects:
                       Estimate     Std. Error         df          t value          Pr(>|t|)
(Intercept)             94.949        2.044           54.032        46.451       <2e-16 ***
targetnanina             3.514        1.989         1061.196         1.767        0.0776 .
```

Table C12. Structure and output of the model testing the effect of TG1 type on S4 duration when boundary is word boundary.

1.2. F0 maximum difference

```
Model: f0 maximum difference
lmer(formula = f0max_diff ~ boundary + (1 | subj), data = dt)
Random effects:
  Groups           Name            Variance       Std.Dev.
  subj           (Intercept)         214.9          14.66
  Residual                          1398.5          37.40
Fixed effects:
                      Estimate        Std. Error           df             t value          Pr(>|t|)
(Intercept)            24.183           2.938             41.324           8.23         3.02e-10 ***
boundarywd            -28.815           1.805           1818.907         -15.96          < 2e-16 ***
```

Table C13. Structure and output of the model testing the effect of boundary on f0 maximum difference.

2. Individual results

2.1. Syllable durations

2.1.1. S1 duration

Participant	Fixed Effects					
1	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	116.73	3.949	29.563	2.00E-16	***
	boundarywd	-6.78	4.517	-1.501	0.138	
	targetnanina	23.108	4.515	5.118	2.80E-06	***
2	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt

	(Intercept)	169.3049	3.8887	43.538	<2e-16	***
	boundarywd	4.9318	4.5146	1.092	0.279	
	targetnanina	0.6576	4.5146	0.146	0.885	
3	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	152.613	2.935	51.989	<2e-16	***
	boundarywd	1.189	3.352	0.355	0.7242	
	targetnanina	-6.698	3.352	-1.998	0.0506	.
4	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	124.4615	3.4983	35.577	<2e-16	***
	boundarywd	-0.6248	4.1961	-0.149	0.8821	
	targetnanina	-9.2575	4.1921	-2.208	0.0309	*
5	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	142.3319	4.3136	32.996	<2e-16	***
	boundarywd	0.3491	5.0543	0.069	0.9452	
	targetnanina	-9.4659	5.0508	-1.874	0.0666	.
6	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	138.4642	5.048	27.43	<2e-16	***
	boundarywd	6.0199	5.703	1.056	0.296	
	targetnanina	-0.8021	5.6212	-0.143	0.887	
7	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	132.937	4.927	26.981	2.00E-16	***
	boundarywd	-4.997	5.369	-0.931	0.35503	
	targetnanina	14.774	5.38	2.746	0.00756	**
8	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	110.2941	3.6575	30.156	<2e-16	***
	boundarywd	8.9177	4.1825	2.132	0.0367	*
	targetnanina	-0.7979	4.1842	-0.191	0.8493	
9	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	136.2461	2.21353	61.551	2.00E-16	***
	boundarywd	-0.09444	2.55597	-0.037	0.971	
	targetnanina	-14.1794	2.55597	-5.548	4.98E-07	***
10	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	130.8	3.013	43.407	<2e-16	***
	boundarywd	2.356	3.48	0.677	0.501	
	targetnanina	3.625	3.48	1.042	0.301	
11	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	97.139	2.921	33.256	<2e-16	***
	boundarywd	7.828	3.354	2.334	0.0231	*
	targetnanina	-2.939	3.357	-0.875	0.385	
12	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	129.841	2.717	47.788	2.00E-16	***

	boundarywd	-2.193	3.168	-0.692	0.491	
	targetnanina	-14.516	3.168	-4.582	2.06E-05	***
13	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	133.22	4.256	31.298	<2e-16	***
	boundarywd	2.232	4.939	0.452	0.653	
	targetnanina	-5.957	4.939	-1.206	0.232	
14	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	113.8	42.46	2.68	0.00923	**
	boundarywd	49.39	49.27	1.002	0.3197	
	targetnanina	50.81	49.27	1.031	0.30606	
15	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	134.778	6.152	21.909	<2e-16	***
	boundarywd	4.023	7.104	0.566	0.5732	
	targetnanina	12.279	7.104	1.729	0.0889	.
16	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	109.043	3.338	32.662	<2e-16	***
	boundarywd	5.831	3.819	1.527	0.1316	
	targetnanina	-6.787	3.819	-1.777	0.0801	.
17	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	135.212	3.746	36.094	<2e-16	***
	boundarywd	8.884	4.326	2.054	0.0438	*
	targetnanina	3.249	4.326	0.751	0.4552	
18	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	111.239	4.766	23.34	<2e-16	***
	boundarywd	5.248	5.479	0.958	0.341	
	targetnanina	5.102	5.479	0.931	0.355	
19	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	111.123	3.004	36.997	<2e-16	***
	boundarywd	-2.469	3.553	-0.695	0.49	
	targetnanina	1.407	3.537	0.398	0.692	
20	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	125.6464	3.7064	33.9	<2e-16	***
	boundarywd	0.1462	4.3003	0.034	0.973	
	targetnanina	14.4289	4.3003	3.355	0.0013	**
21	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	152.087	3.283	46.327	<2e-16	***
	boundarywd	5.117	3.791	1.35	0.181	
	targetnanina	-1.851	3.793	-0.488	0.627	
22	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	115.564	3.438	33.617	<2e-16	***
	boundarywd	-2.054	3.95	-0.52	0.605	

	targetnanina	4.394	3.953	1.112	0.27	
23	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	112.0208	2.4566	45.6	<2e-16	***
	boundarywd	0.08333	2.83664	0.029	0.977	
	targetnanina	-3.10556	2.83664	-1.095	0.277	
24	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	125.412	2.968	42.256	2.00E-16	***
	boundarywd	-6.291	3.427	-1.836	0.0707	.
	targetnanina	14.379	3.427	4.196	7.96E-05	***
25	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	164.1151	3.0468	53.865	<2e-16	***
	boundarywd	-9.2292	3.5181	-2.623	0.0107	*
	targetnanina	0.7997	3.5181	0.227	0.8209	
26	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	143.076	4.785	29.901	2.00E-16	***
	boundarywd	15.378	5.525	2.783	0.00694	**
	targetnanina	8.79	5.525	1.591	0.11622	
27	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	99.618	2.806	35.501	<2e-16	***
	boundarywd	1.821	3.271	0.557	0.5796	
	targetnanina	-7.376	3.272	-2.254	0.0275	*
28	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	123.589	3.97	31.13	2.00E-16	***
	boundarywd	-6.447	4.584	-1.406	0.164101	
	targetnanina	17.037	4.584	3.716	0.000407	***
29	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	122.6555	5.3683	22.848	<2e-16	***
	boundarywd	0.9603	6.1755	0.155	0.877	
	targetnanina	-4.0492	6.1696	-0.656	0.514	
30	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	109.315	4.136	26.427	<2e-16	***
	boundarywd	-6.716	4.85	-1.385	0.1708	
	targetnanina	10.063	4.85	2.075	0.0419	*
31	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	123.9291	3.46032	35.814	<2e-16	***
	boundarywd	-0.05535	4.01479	-0.014	0.989	
	targetnanina	-2.68647	4.01479	-0.669	0.506	
32	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	115.7888	3.11415	37.182	2.00E-16	***
	boundarywd	-0.07219	3.54561	-0.02	0.98382	

	targetnanina	-12.1594	3.54561	-3.429	0.00103	**

Table C14. Structure and output of the individual models testing the effect of boundary on S1 duration.

2.1.2. S2 duration

Participant	Fixed Effects					
1	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	186.946	3.586	52.135	2.00E-16	***
	boundarywd	-19.875	4.102	-4.845	7.84E-06	***
	targetnanina	-7.759	4.101	-1.892	0.0628	.
2	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	190.104	2.684	70.838	2.00E-16	***
	boundarywd	-16.407	3.116	-5.266	1.98E-06	***
	targetnanina	-3.885	3.116	-1.247	0.217	
3	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	176.293	2.378	74.149	2.00E-16	***
	boundarywd	-9.125	2.715	-3.361	0.00142	**
	targetnanina	2.684	2.715	0.989	0.32718	
4	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	153.033	2.767	55.316	<2e-16	***
	boundarywd	-3.655	3.318	-1.102	0.2749	
	targetnanina	7.173	3.315	2.164	0.0344	*
5	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	165.108	3.379	48.86	<2e-16	***
	boundarywd	-1.741	3.959	-0.44	0.662	
	targetnanina	-4.944	3.957	-1.25	0.217	
6	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	157.582	3.572	44.115	<2e-16	***
	boundarywd	6.406	4.036	1.587	0.1187	
	targetnanina	-10.445	3.978	-2.626	0.0114	*
7	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	169.785	4.816	35.255	<2e-16	***
	boundarywd	6.055	5.248	1.154	0.2523	
	targetnanina	10.04	5.259	1.909	0.0601	.
8	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	159.538	3.388	47.089	2.00E-16	***
	boundarywd	5.292	3.874	1.366	0.17656	
	targetnanina	-10.285	3.876	-2.654	0.00994	**
9	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt

	(Intercept)	167.087	3.498	47.762	2.00E-16	***
	boundarywd	2.631	4.04	0.651	0.517034	
	targetnanina	-14.095	4.04	-3.489	0.000848	***
10	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	202.916	3.332	60.906	<2e-16	***
	boundarywd	-5.106	3.847	-1.327	0.1888	
	targetnanina	-7.295	3.847	-1.896	0.0621	.
11	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	150.44	3.135	47.992	<2e-16	***
	boundarywd	2.308	3.599	0.641	0.524	
	targetnanina	-1.993	3.603	-0.553	0.582	
12	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	172.801	2.976	58.057	2.00E-16	***
	boundarywd	-13.856	3.471	-3.992	0.000165	***
	targetnanina	8.873	3.471	2.557	0.01284	*
13	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	147.851	2.909	50.821	2.00E-16	***
	boundarywd	-6.161	3.375	-1.825	0.07237	.
	targetnanina	-11.861	3.375	-3.514	0.00079	***
14	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	189.448	3.155	60.047	2.00E-16	***
	boundarywd	-23.443	3.661	-6.404	1.65E-08	***
	targetnanina	6.048	3.661	1.652	0.103	
15	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	168.152	3.417	49.208	<2e-16	***
	boundarywd	4.958	3.946	1.257	0.214	
	targetnanina	-1.191	3.946	-0.302	0.764	
16	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	158.543	2.682	59.123	2.00E-16	***
	boundarywd	-11.546	3.068	-3.764	0.000355	***
	targetnanina	-3.972	3.068	-1.295	0.199859	
17	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	178.451	2.646	67.447	<2e-16	***
	boundarywd	-4.514	3.055	-1.478	0.144	
	targetnanina	-3.906	3.055	-1.278	0.205	
18	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	177.887	4.615	38.543	<2e-16	***
	boundarywd	-10.626	5.305	-2.003	0.0491	*
	targetnanina	4.225	5.305	0.796	0.4285	
19	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	165.514	3.397	48.718	2.00E-16	***

	boundarywd	-24.323	4.018	-6.053	9.52E-08	***
	targetnanina	-1.764	4.001	-0.441	0.661	
20	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	183.202	3.247	56.416	2.00E-16	***
	boundarywd	7.996	3.768	2.122	0.0375	*
	targetnanina	-18.954	3.768	-5.031	3.80E-06	***
21	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	144.916	1.832	79.093	2.00E-16	***
	boundarywd	-6.364	2.116	-3.007	0.00367	**
	targetnanina	-2.611	2.117	-1.233	0.22163	
22	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	159.004	2.86	55.587	2.00E-16	***
	boundarywd	-19.191	3.286	-5.84	1.75E-07	***
	targetnanina	9.155	3.289	2.783	0.00701	**
23	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	173.551	2.362	73.491	2.00E-16	***
	boundarywd	4.919	2.727	1.804	0.0756	.
	targetnanina	-14.429	2.727	-5.291	1.36E-06	***
24	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	180.531	2.712	66.562	<2e-16	***
	boundarywd	-8.079	3.132	-2.58	0.012	*
	targetnanina	-3.902	3.132	-1.246	0.217	
25	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	193.981	2.511	77.264	2.00E-16	***
	boundarywd	-11.075	2.899	-3.82	0.000288	***
	targetnanina	-14.956	2.899	-5.159	2.27E-06	***
26	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	181.829	3.134	58.015	<2e-16	***
	boundarywd	-6.655	3.619	-1.839	0.0702	.
	targetnanina	3.396	3.619	0.938	0.3513	
27	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	157.1115	1.90775	82.354	2.00E-16	***
	boundarywd	-10.0191	2.22356	-4.506	2.72E-05	***
	targetnanina	0.06806	2.22447	0.031	0.976	
28	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	159.916	2.21	72.365	<2e-16	***
	boundarywd	-2.329	2.552	-0.913	0.365	
	targetnanina	3.695	2.552	1.448	0.152	
29	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	169.411	4.96	34.154	<2e-16	***
	boundarywd	-14.66	5.706	-2.569	0.0126	*

	targetnanina	-3.432	5.701	-0.602	0.5493	
32	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	151.217	2.525	59.884	<2e-16	***
	boundarywd	-40.268	2.961	-13.599	<2e-16	***
	targetnanina	5.095	2.961	1.721	0.09	.
31	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	183.389	3.971	46.184	2.00E-16	***
	boundarywd	-18.053	4.607	-3.919	0.000209	***
	targetnanina	-24.41	4.607	-5.298	1.36E-06	***
32	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	156.108	2.642	59.09	<2e-16	***
	boundarywd	-0.773	3.008	-0.257	0.798	
	targetnanina	6.142	3.008	2.042	0.045	*

Table C15. Structure and output of the individual models testing the effect of boundary on S2 duration.

2.1.3. S3 duration

Participant	Fixed Effects					
1	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	270.522	5.933	45.593	<2e-16	***
	boundarywd	-125.207	6.788	-18.446	<2e-16	***
	targetnanina	-10.9	6.785	-1.606	0.113	
2	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	277.539	5.848	47.46	<2e-16	***
	boundarywd	-96.535	6.789	-14.219	<2e-16	***
	targetnanina	-15.864	6.789	-2.337	0.0228	*
3	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	249.835	4.851	51.502	2.00E-16	***
	boundarywd	-76.578	5.539	-13.825	2.00E-16	***
	targetnanina	-25.443	5.539	-4.593	2.58E-05	***
4	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	221.466	4.229	52.369	2.00E-16	***
	boundarywd	-73.881	5.072	-14.565	2.00E-16	***
	targetnanina	-16.276	5.068	-3.212	0.00209	**
5	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	231.356	6.307	36.683	< 2e-16	***
	boundarywd	-65.382	7.39	-8.848	7.06E-12	***
	targetnanina	-23.912	7.385	-3.238	0.00212	**
6	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt

	(Intercept)	213.79	7.4	28.892	2.00E-16	***
	boundarywd	-53.48	8.36	-6.397	5.27E-08	***
	targetnanina	-16.3	8.24	-1.978	0.0535	.
7	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	241.284	5.513	43.762	<2e-16	***
	boundarywd	-68.06	6.008	-11.328	<2e-16	***
	targetnanina	-17.887	6.02	-2.971	0.004	**
8	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	263.692	5.737	45.963	2.00E-16	***
	boundarywd	-106.262	6.561	-16.197	2.00E-16	***
	targetnanina	-22.065	6.563	-3.362	0.00128	**
9	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	219.701	4.589	47.87	2.00E-16	***
	boundarywd	-73.32	5.299	-13.84	2.00E-16	***
	targetnanina	-28.776	5.299	-5.43	7.92E-07	***
10	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	245.76	3.94	62.373	2.00E-16	***
	boundarywd	-97.95	4.55	-21.53	2.00E-16	***
	targetnanina	-12.19	4.55	-2.679	0.00922	**
11	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	219.901	3.508	62.687	<2e-16	***
	boundarywd	-78.172	4.028	-19.409	<2e-16	***
	targetnanina	-9.091	4.032	-2.255	0.0279	*
12	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	230.546	3.161	72.934	2.00E-16	***
	boundarywd	-89.991	3.686	-24.416	2.00E-16	***
	targetnanina	-10.737	3.686	-2.913	0.00486	**
13	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	256.645	4.088	62.786	<2e-16	***
	boundarywd	-130.374	4.743	-27.49	<2e-16	***
	targetnanina	-8.439	4.743	-1.779	0.0796	.
14	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	275.385	5.767	47.756	<2e-16	***
	boundarywd	-130.066	6.691	-19.44	<2e-16	***
	targetnanina	-7.729	6.691	-1.155	0.252	
15	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	239.567	5.3	45.204	<2e-16	***
	boundarywd	-76.164	6.12	-12.446	<2e-16	***
	targetnanina	-9.142	6.12	-1.494	0.14	
16	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	215.155	5.546	38.797	<2e-16	***

	boundarywd	-90.245	6.344	-14.225	<2e-16	***
	targetnanina	-3.702	6.344	-0.583	0.562	
17	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	286.743	4.407	65.059	<2e-16	***
	boundarywd	-134.129	5.089	-26.355	<2e-16	***
	targetnanina	-10.563	5.089	-2.076	0.0417	*
18	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	249.782	4.547	54.93	<2e-16	***
	boundarywd	-107.76	5.227	-20.615	<2e-16	***
	targetnanina	-8.6	5.227	-1.645	0.104	
19	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	262.982	6.769	38.852	2.00E-16	***
	boundarywd	-107.497	8.006	-13.427	2.00E-16	***
	targetnanina	-27.619	7.971	-3.465	0.00098	***
20	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	257.942	3.349	77.017	<2e-16	***
	boundarywd	-95.113	3.886	-24.477	<2e-16	***
	targetnanina	-5.134	3.886	-1.321	0.191	
21	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	247.639	4.773	51.883	2.00E-16	***
	boundarywd	-100.425	5.512	-18.219	2.00E-16	***
	targetnanina	-14.64	5.514	-2.655	0.00984	**
22	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	259.319	2.952	87.835	2.00E-16	***
	boundarywd	-124.566	3.392	-36.724	2.00E-16	***
	targetnanina	-19.096	3.395	-5.625	4.07E-07	***
23	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	215.117	4.043	53.213	<2e-16	***
	boundarywd	-81.009	4.668	-17.354	<2e-16	***
	targetnanina	-14.924	4.668	-3.197	0.0021	**
24	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	203.708	4.084	49.875	2.00E-16	***
	boundarywd	-58.959	4.716	-12.501	2.00E-16	***
	targetnanina	-14.91	4.716	-3.161	0.00233	**
25	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	236.351	5.79	40.824	2.00E-16	***
	boundarywd	-67.174	6.685	-10.048	3.83E-15	***
	targetnanina	-18.151	6.685	-2.715	0.00836	**
26	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	236.002	3.553	66.425	2.00E-16	***
	boundarywd	-67.752	4.103	-16.515	2.00E-16	***

	targetnanina	-23.692	4.103	-5.775	2.02E-07	***
27	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	227.968	3.742	60.924	2.00E-16	***
	boundarywd	-102.627	4.361	-23.531	2.00E-16	***
	targetnanina	-21.557	4.363	-4.941	5.47E-06	***
28	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	280.587	5.864	47.845	<2e-16	***
	boundarywd	-137.027	6.772	-20.235	<2e-16	***
	targetnanina	-14.879	6.772	-2.197	0.0314	*
29	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	213.368	4.581	46.576	<2e-16	***
	boundarywd	-73.163	5.27	-13.883	<2e-16	***
	targetnanina	-9.939	5.265	-1.888	0.0637	.
30	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	234.576	4.206	55.774	<2e-16	***
	boundarywd	-115.347	4.932	-23.389	<2e-16	***
	targetnanina	-7.479	4.932	-1.516	0.134	
31	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	175.022	4.167	41.998	2.00E-16	***
	boundarywd	-18.155	4.835	-3.755	0.00036	***
	targetnanina	-25.783	4.835	-5.332	1.19E-06	***
32	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	311.483	4.635	67.205	2.00E-16	***
	boundarywd	-156.599	5.277	-29.676	2.00E-16	***
	targetnanina	-25.626	5.277	-4.856	7.35E-06	***

Table C16. Structure and output of the individual models testing the effect of boundary on S3 duration.

2.1.4. S4 duration

Participant	Fixed Effects					
1	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	91.7326	5.0765	18.07	<2e-16	***
	boundarywd	-4.4119	5.8076	-0.76	0.45	
	targetnanina	0.0312	5.8052	0.005	0.996	
2	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	76.704	6.826	11.236	2.00E-16	***
	boundarywd	25.668	7.925	3.239	0.00196	**
	targetnanina	18.175	7.925	2.293	0.02535	*
3	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt

	(Intercept)	68.484	6.532	10.485	9.95E-15	***
	boundarywd	33.251	7.458	4.458	4.11E-05	***
	targetnanina	9.72	7.458	1.303	0.198	
4	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	85.457	4.685	18.239	<2e-16	***
	boundarywd	2.652	5.62	0.472	0.639	
	targetnanina	3.16	5.615	0.563	0.576	
5	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	61.451	4.931	12.462	< 2e-16	***
	boundarywd	28.126	5.778	4.868	1.13E-05	***
	targetnanina	14.366	5.774	2.488	0.0162	*
6	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	82.341	8.914	9.237	2.17E-12	***
	boundarywd	16.001	10.071	1.589	0.118	
	targetnanina	10.535	9.927	1.061	0.294	
7	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	87.229	7.586	11.499	<2e-16	***
	boundarywd	14.76	8.266	1.786	0.0783	.
	targetnanina	11.576	8.283	1.398	0.1664	
8	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	64.01	5.568	11.497	<2e-16	***
	boundarywd	16.56	6.367	2.601	0.0114	*
	targetnanina	15.669	6.369	2.46	0.0165	*
9	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	58.31	5.428	10.743	2.24E-16	***
	boundarywd	22.67	6.267	3.617	0.00056	***
	targetnanina	15.54	6.267	2.48	0.0156	*
10	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	85.038	5.88	14.463	2.00E-16	***
	boundarywd	25.493	6.789	3.755	0.00036	***
	targetnanina	-6.971	6.789	-1.027	0.30814	
11	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	75.474	5.1	14.8	2.00E-16	***
	boundarywd	20.367	5.855	3.479	0.00096	***
	targetnanina	-3.906	5.861	-0.666	0.50777	
12	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	68.439	4.991	13.713	2.00E-16	***
	boundarywd	47.989	5.819	8.247	8.65E-12	***
	targetnanina	-1.382	5.819	-0.237	0.813	
13	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	48.58	41.14	1.181	0.242	

	boundarywd	76.51	47.74	1.603	0.114	
	targetnanina	42.14	47.74	0.883	0.38	
14	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	125.07	6.413	19.502	2.00E-16	***
	boundarywd	-4.93	7.441	-0.663	0.50985	
	targetnanina	-23.125	7.441	-3.108	0.00275	**
15	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	78.131	6.942	11.254	2.00E-16	***
	boundarywd	31.209	8.016	3.893	0.00025	***
	targetnanina	3.329	8.016	0.415	0.67936	
16	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	59.159	4.705	12.574	<2e-16	***
	boundarywd	14.436	5.382	2.682	0.0092	**
	targetnanina	18.366	5.382	3.412	0.0011	**
17	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	62.218	5.048	12.326	2.00E-16	***
	boundarywd	34.798	5.829	5.97	9.20E-08	***
	targetnanina	15.96	5.829	2.738	0.00785	**
18	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	56.168	5.082	11.051	2.00E-16	***
	boundarywd	28.671	5.842	4.907	5.80E-06	***
	targetnanina	11.545	5.842	1.976	0.0521	.
19	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	70.475	5.898	11.948	<2e-16	***
	boundarywd	11.097	6.976	1.591	0.1169	
	targetnanina	13.697	6.946	1.972	0.0531	.
20	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	78.294	6.242	12.544	<2e-16	***
	boundarywd	22.741	7.242	3.14	0.0025	**
	targetnanina	9.88	7.242	1.364	0.177	
21	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	56.244	6.205	9.065	2.28E-13	***
	boundarywd	27.071	7.166	3.778	0.00033	***
	targetnanina	4.313	7.168	0.602	0.54937	
22	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	58.998	5.336	11.056	2.00E-16	***
	boundarywd	24.14	6.131	3.937	0.0002	***
	targetnanina	7.129	6.136	1.162	0.24946	
23	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	59.017	4.986	11.836	2.00E-16	***
	boundarywd	19.716	5.757	3.424	0.00104	**

	targetnanina	9.914	5.757	1.722	0.08958	.
24	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	106.034	6.842	15.496	<2e-16	***
	boundarywd	-9.086	7.901	-1.15	0.254	
	targetnanina	-1.772	7.901	-0.224	0.823	
25	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	98.859	7.668	12.892	<2e-16	***
	boundarywd	10.454	8.854	1.181	0.242	
	targetnanina	3.867	8.854	0.437	0.664	
26	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	81.617	5.667	14.403	<2e-16	***
	boundarywd	13.638	6.543	2.084	0.0408	*
	targetnanina	7.039	6.543	1.076	0.2858	
27	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	44.554	4.519	9.859	1.14E-14	***
	boundarywd	29.156	5.267	5.535	5.59E-07	***
	targetnanina	15.974	5.269	3.031	0.00346	**
28	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	87.349	7.638	11.436	<2e-16	***
	boundarywd	5.029	8.82	0.57	0.57	
	targetnanina	4.069	8.82	0.461	0.646	
29	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	67.795	5.099	13.296	2.00E-16	***
	boundarywd	16.091	5.866	2.743	0.00794	**
	targetnanina	6.596	5.86	1.126	0.26471	
30	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	83.71	4.456	18.786	<2e-16	***
	boundarywd	13.687	5.225	2.62	0.0109	*
	targetnanina	1.572	5.225	0.301	0.7645	
31	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	64.503	4.656	13.854	2.00E-16	***
	boundarywd	22.227	5.402	4.115	0.00011	***
	targetnanina	1.611	5.402	0.298	0.76642	
32	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	58.551	4.614	12.69	2.00E-16	***
	boundarywd	27.856	5.253	5.303	1.34E-06	***
	targetnanina	11.314	5.253	2.154	0.0348	*

Table C17. Structure and output of the individual models testing the effect of boundary on S4 duration.

2.2. f0 maximum difference

Participant	Fixed Effects					
1	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	12.918	4.212	3.067	0.0031	**
	boundarywd	-11.326	5.873	-1.928	0.058	.
2	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	31.526	5.434	5.801	2.52E-07	***
	boundarywd	-35.396	7.747	-4.569	2.44E-05	***
3	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	5.904	4.563	1.294	0.201	
	boundarywd	-8.855	6.453	-1.372	0.175	
4	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	18.602	1.196	15.55	<2e-16	***
	boundarywd	-17.604	1.749	-10.06	1.68E-14	***
5	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	-16.433	5.831	-2.818	0.0069	**
	boundarywd	10.932	7.947	1.376	0.175	
6	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	10.858	7.166	1.515	0.136	
	boundarywd	-16.092	9.28	-1.734	0.0891	.
7	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	16.57	3.58	4.628	1.52E-05	***
	boundarywd	-17.317	4.907	-3.529	0.000716	***
8	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	0.1868	9.6083	0.019	0.985	
	boundarywd	-17.6917	17.1542	-1.031	0.307	
9	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	33.746	4.488	7.519	1.51E-10	***
	boundarywd	-39.089	6.303	-6.201	3.60E-08	***
10	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	15.0536	6.0295	2.497	0.0149	*
	boundarywd	0.4252	8.4676	0.05	0.9601	
11	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	0.9633	1.4026	0.687	0.495	
	boundarywd	-1.0842	2.0001	-0.542	0.59	
12	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	56.396	5.548	10.166	2.78E-15	***
	boundarywd	-70.084	7.845	-8.933	4.46E-13	***
13	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt

	(Intercept)	45.057	9.491	4.748	1.10E-05	***
	boundarywd	-49.36	13.618	-3.625	0.000554	***
14	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	0.7488	5.9802	0.125	0.9007	
	boundarywd	-17.3628	8.5175	-2.038	0.0453	*
15	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	3.459	14.648	0.236	0.814	
	boundarywd	16.464	20.523	0.802	0.426	
16	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	62.176	2.006	31	<2e-16	***
	boundarywd	-80.84	2.754	-29.35	<2e-16	***
17	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	56.04	15.93	3.518	0.00106	**
	boundarywd	-84.26	18.12	-4.65	3.29E-05	***
18	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	40.44	9.5	4.257	0.000101	***
	boundarywd	-58.76	11.29	-5.206	4.40E-06	***
19	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	4.462	6.748	0.661	0.511	
	boundarywd	-8.599	9.707	-0.886	0.379	
20	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	12.229	8.843	1.383	0.173	
	boundarywd	-26.862	10.675	-2.516	0.0152	*
21	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	131.457	9.244	14.221	<2e-16	***
	boundarywd	-114.794	13.073	-8.781	8.39E-13	***
22	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	17.272	3.76	4.594	2.19E-05	***
	boundarywd	-15.442	5.159	-2.993	0.00396	**
23	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	36.866	4.19	8.799	6.93E-13	***
	boundarywd	-44.936	5.884	-7.637	9.21E-11	***
24	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	37.439	1.846	20.28	<2e-16	***
	boundarywd	-37.519	2.24	-16.75	<2e-16	***
25	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	12.649	1.208	10.474	8.32E-15	***
	boundarywd	-13.435	1.652	-8.132	4.68E-11	***
26	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	15.31	16.35	0.937	0.356	
	boundarywd	-18.2	17.75	-1.025	0.313	

27	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	33.59	15.28	2.198	0.0347	*
	boundarywd	-41.01	15.72	-2.61	0.0133	*
28	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	87.047	8.536	10.198	6.59E-14	***
	boundarywd	-92.443	10.357	-8.925	5.36E-12	***
29	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	2.084	23.708	0.088	0.931	
	boundarywd	28.101	29.57	0.95	0.361	
30	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	24.35	15.84	1.537	0.1344	
	boundarywd	-36.88	16.34	-2.257	0.0312	*
31	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	-8.294	5.959	-1.392	0.169	
	boundarywd	4.033	8.047	0.501	0.618	
32	Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	sig.annt
	(Intercept)	-9.293	1.057	-8.792	2.54E-12	***
	boundarywd	11.607	1.395	8.318	1.59E-11	***

Table C18. Structure and output of the individual models testing the effect of boundary on f0 maximum difference.

APPENDIX D

Figures Representing the Boundary-related Effect of TG1 TYPE on Syllable Durations for Individual Speakers

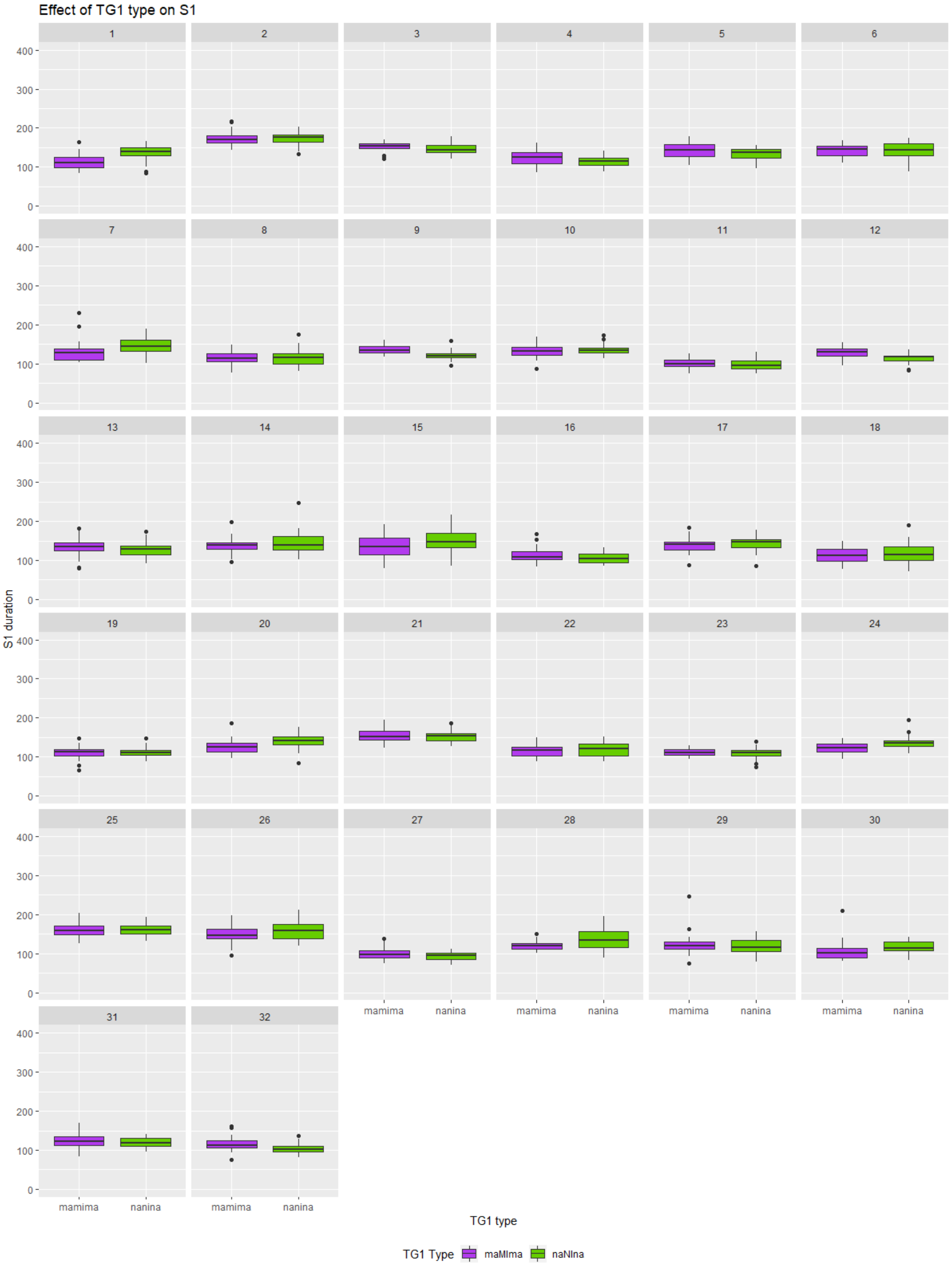


Figure D.1. Effect of TG1 type on S1 duration for 32 individual participants.

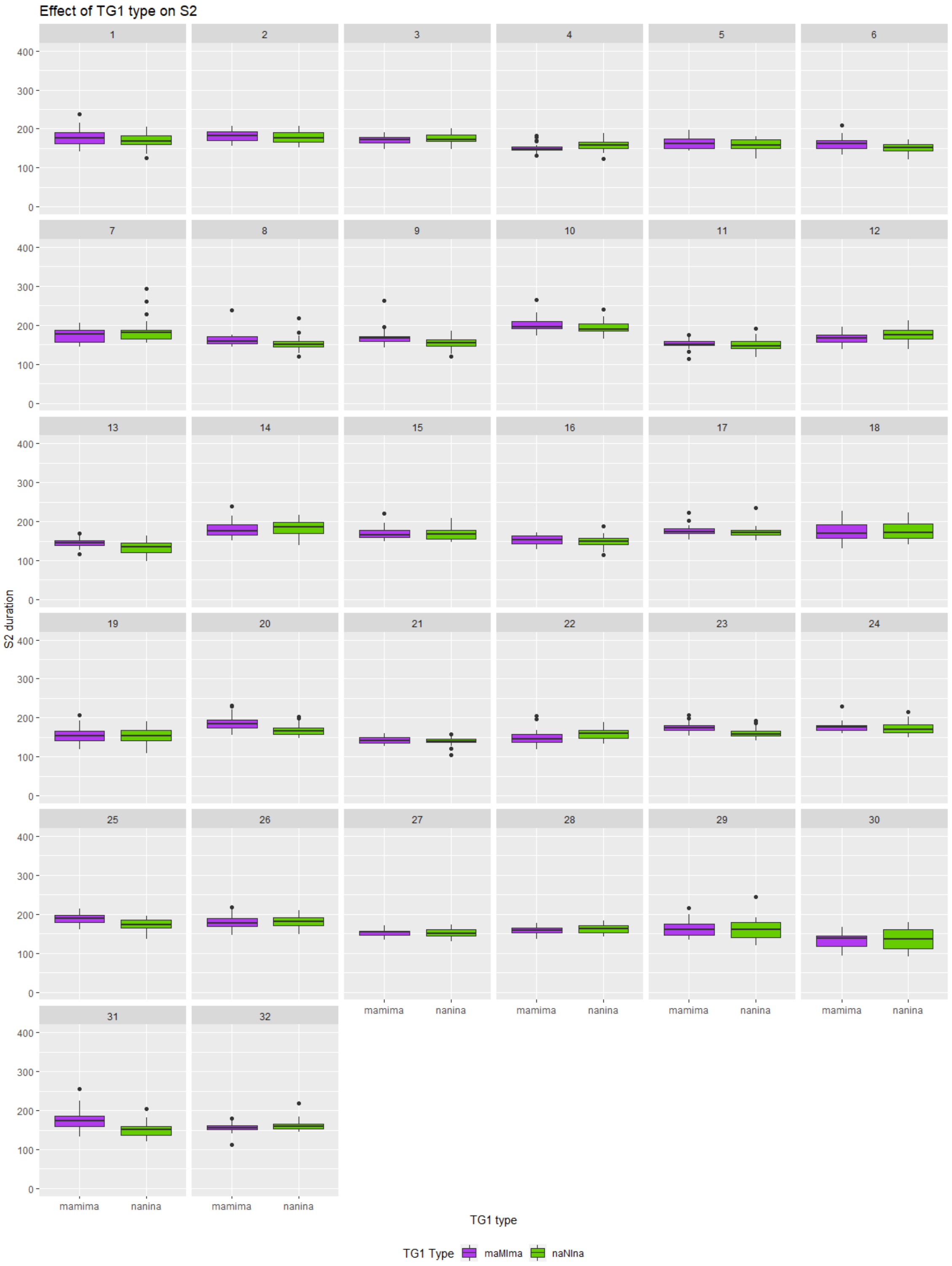


Figure D.2. Effect of TG1 type on S2 duration for 32 individual participants.

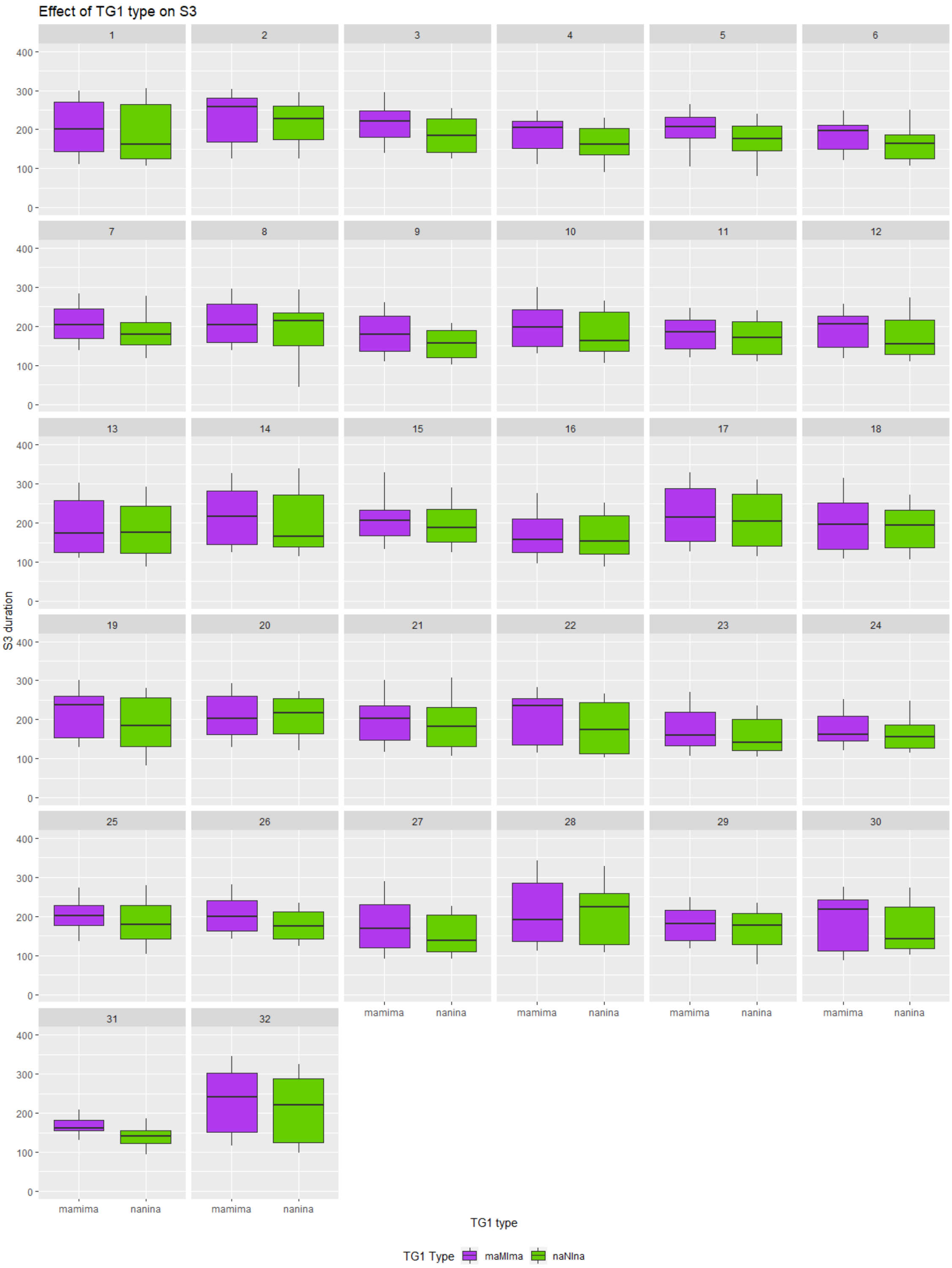


Figure D.3. Effect of TG1 type on S3 duration for 32 individual participants.

Figure D.4. Effect of TG1 type on S4 duration for 32 individual participants.

APPENDIX E

Figure Representing the Significant Effect of TG1 TYPE on f0 Maximum Difference (Δf0; in Hz) for Individual Speakers

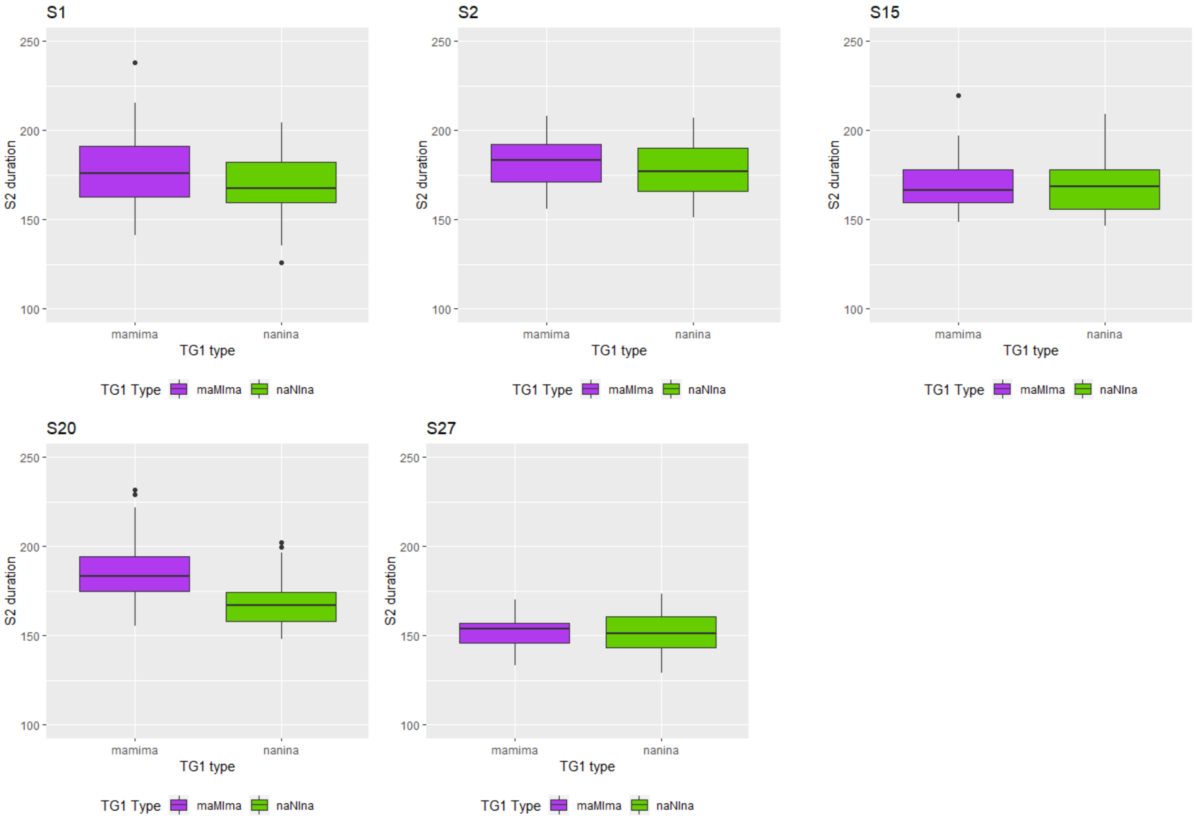


Figure E.1. Effect of TG1 Type on f0 maximum difference in five participants.

APPENDIX F

Visual Stimuli for the Perception Study

Figure F1. Visual Stimuli for the Perception Study.

BIBLIOGRAPHY

Beach, C. (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.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. Phonology Yearbook, 3, 255-309.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In Jun, S-A (ed.), *Prosodic Typology*: *The Phonology of Intonation and Phrasing* (pp. 9-54). Oxford: Oxford University Press.

Beddor, P. S. (2009). A coarticulatory path to sound change. Language, 85, 785-821.

Beddor, P. S., McGowan, K. B., Boland, J., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. Journal of the Acoustical Society of America 133(4), 2350-2366.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. Language, 94 (4), 931-968.

Bent, T., & Holt, R. F. (2017). Representation of speech variability. Wiley Interdisciplinary Reviews: Cognitive Science, 1–14. doi: 10.1002/wcs.1434

Berkovits, R. (1993). Utterance-final lengthening and the duration of final-stop closures. Journal of Phonetics, 21, 479-489.

Berkovits, R. (1994). Durational effects in final lengthening, gapping, and contrastive stress. Language and Speech, 37, 237-250.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.14, retrieved 2 May 2020 from http://www.praat.org/

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. Journal of Phonetics, 52, 46-57.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D.J., & Kerkhofs, R., (2010). The interplay between prosody and syntax in sentence processing: the case of subject- and object-control verbs. Journal of Cognitive Neuroscience, 22 (5), 1036-1053.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. Phonetica, 49, 155-180.

Browman, C.P., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R.F. Port, & T. van Gelder (eds.), *Mind as Motion* (pp. 175-193). Cambridge, MA: MIT Press.

Byrd, D., & Riggs, D. (2008). Locality interactions with prominence in determining the scope of phrasal lengthening. Journal of International Phonetic Association, 38(2), 187-202.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple phrasal boundaries. Journal of Phonetics, 26, 173-199.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. Journal of Phonetics, 31(2), 149-180.

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of phrase boundary effects. Journal of the Acoustical Society of America, 120, 1589-1599.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. Phonetica, 57(1), 3-16.

Cambier-Langeveld, T., Nespor, M., & van Heuven, V. J. (1997). The domain of final lengthening in production and perception in Dutch. In Proceedings of Eurospeech-1997, 931-934.

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific production in intonation. In S. Fuchs, D. Pape, C. Petrone & P. Perrier (eds.), *Individual Differences in Speech Production and Perception* (pp. 123-145). Frankfurt am Main: Peter Lang.

Cho, H., & Hirst, D. J. (2006). The contribution of silent pauses to the perception of prosodic boundaries in Korean read speech. In Proceedings of Speech Prosody 2006, Dresden, Germany.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. Journal of Phonetics, 29, 155-190.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. Journal of Phonetics, 37, 466-485.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. Journal of Phonetics, 66, 185-216.

Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: what do listeners imitate?. In Proceedings of Interspeech 2011, Florence.

Cole, J. Mo, Y., & Baek, S. (2010a). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. Language and Cognitive Processes, 25: 7, 1141 – 1177.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010b). Signal-based and expectation-based factors in the perception of prosodic prominence. Laboratory Phonology, 110(2), 425-452.

Cole, J. (2015). Prosody in context: a review. Language Cognition and Neuroscience, 30(1-2), 1-31.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (eds). *Laboratory Phonology 10: Variability, Phonetic Detail and Phonological Representation* (pp. 91-111). Berlin: de Gruyter.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes, 16, 507-534.

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. Journal of the Acoustical Society of America, 96, 2037-2047.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. Journal of Memory and Language, 59, 294-311.

Dilley, L. & Shattuck-Hufnagel, S. (1995). Variability in glottalization of word onset vowels in American English. In Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS), Stockholm, 586-589.

Edwards, J. E., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. Journal of the Acoustical Society of America, 89, 369-382.

Escudero, P., Benders, T., & Lipski, S. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. Journal of Phonetics, 37, 452-466.

Fant, G., Kruckenberg, A., & Ferreira, J. B. (2003). Individual variations in pausing. A study of read speech. In Proceedings of PHONUM 9, 193-196.

Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment?. BMC medical research methodology, 2, 8. DOI: 10.1186/1471-2288-2-8.

Ferreira, F. (1993). Creating of prosody during sentence production. Psychological Review, 100(2), 233-253.

Fougeron, C., & P. Keating (1997). Articulatory strengthening at edges of prosodic domains. Journal of the Acoustical Society of America, 101, 3728-3740.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist approach. Journal of Phonetics, 14, 3-28.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1166-1183.

Goldinger, S. D. (1997). Words and voices: Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (eds.), *Talker variability in speech perception* (pp. 33-66). San Diego, CA: Academic Press.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. Arbib (ed.), *Action to Language via the Mirror Neuron System* (pp. 215-249). New York, NY: Cambridge University Press.

Gollrad, A., Sommerfeld, E., & Kügler, F. (2010). Prosodic cue weighting in disambiguation: case ambiguity in German, In Proceedings of Speech Prosody 2010, Paper 165.

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. Journal of Phonetics, 64, 90-107.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. Linguistics, 21, 501-529.

Grosvald, M., & Corina, D. (2012). The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In M.J. Solé, M.J. & D. Recasens (eds.), *The initiation of sound change: Production, perception, and social factors* (pp. 77-100). Amsterdam: John Benjamins.

Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. Journal of Phonetics, 37, 173-88.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. Journal of the Acoustical Society of America, 123, 2825–2835.

Holzgrefe, J., Wellmann, C., Petrone, C., Truckenbrodt, H., Hohle, B., & Wartenburger, I. (2013). Brain response to prosodic boundary cues depends on boundary position. Frontiers in Psychology, 4, Article 421.

Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Räling, R., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2016) How pitch change and final lengthening cue boundary perception in German: converging evience from ERPs and prosodic judgements. Language, Cognition and Neuroscience, 31(7), 904-920.

Horne, M., Strangert, E., & Heldner, M. (1995). Prosodic boundary strength in Swedish: Final lengthening and silent interval duration. In Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS), Stockholm, 170-173.

Hsu, C-S., & Jun, S-A. (1998). Prosodic Strengthening in Taiwanese: Syntagmatic or Paradigmatic?, UCLA Working Papers in Phonetics, 96, 69-89.

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. Journal of the Acoustical Society of America, 132, 3950-3964.

Ito, K., & Speer, S. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. Journal of Memory and Language, 58, 541-573.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. Journal of the Acoustical Society of America, 94, 701–714.

Kataoka, R. (2011). Phonetic and cognitive bases of sound change. Berkeley: University of California, Berkeley dissertation.

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. Journal of Phonetics, 55, 149-181.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014)., The coordination of boundary tones and their interaction with prominence. Journal of Phonetics, 44, 62-82.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2004). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (eds.), *Phonetic Interpretation (Papers in Laboratory Phonology VI)* (pp. 143-161). Cambridge: Cambridge University Press.

Kim, D., Clayards, M., & Goad, H. (2017). Individual differences in second language speech perception across tasks and contrasts: The case of English vowel contrasts by Korean learners. In: Linguistics Vanguard. DOI:10.1515/lingvan-2016-0025.

Klatt, D. H. (1975). Voice-onset time, frication, and aspiration in word-initial consonant clusters. Journal of Speech and Hearing Research, 18, 686-706.

Kong, E., & Edward, J. (2011). Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking. In Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS), Hong Kong, 1126-1129.

Kong, E. J. & Edwards, J. (2016). Individual differences in categorical perception of speech: cue weighting and executive function. Journal of Phonetics, 59, 40-57.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. Journal of Phonetics, 35, 162-179.

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. Journal of Phonetics, 71, 194-214.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn’t it - Listeners make rapid use of prosody to infer speaker meanings. Cognition, 133, 335-342.

Ladd, D. R. (1988). Declination “reset” and the hierarchical organization of utterances. Journal of the Acoustical Society of America, 84, 530-544.

Ladd, D. R. (2008). Intonational phonology. 2nd edition. Cambridge: Cambridge University Press.

Lee, E-K., Mo, Y., & Watson, D. (2008). Eye-tracking evidence for the active role of intonational boundaries in parsing decisions. Poster presented at CUNY 2008: Conference on Human Sentence Processing, Chapel Hill.

Lehiste, I. (1980) Phonetic manifestation of syntactic structure in English. Annual Bulletin Research Institute of Logopedics and Phoniatrics, 14, 1-27.

Lehiste, I., Olive, J. P., & Streeter, L.A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. Journal of the Acoustical Society of America, 60, 1199-1202.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. Glossa, 7, 107-123.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech revised. Cognition, 21, 1-36.

Lin, H-Y., & Fon, J. (2009). Perception of temporal cues at discourse boundaries. In Proceedings of Interspeech-2009, 808-811.

Männel, C., & Friederici, A. D. (2009). Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. Journal of Cognitive Neuroscience, 21, 1988-2006.

Männel, C., & Friederici, A. D. (2016). Neural correlates of prosodic boundary perception in German preschoolers: if pause is present, pitch can go. Brain Research, 1632, 27-33.

Männel, C., Schipke, C., & Friederici, A. D. (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.

Mo, Y. & Cole, J. (2010). Modeling perceived prosody – speaker-dependent vs. speaker-independent models. Poster presented at LabPhon 12, Albuquerque.

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngral articulation - Is it mediated by accentuation?. Journal of Phonetics, 44, 47-61.

Nespor, M., & Vogel, I. (1986). Prosodic phonology. Foris Publications.

Niebuhr, O., D'Imperio, M., Gili Fivela, B., & Cangemi, F. (2011). Are there "shapers" and "aligners"? Individual differences in signalling pitch accent category. In Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS), Hong Kong, 120-123.

Norris D., McQueen J. M., & Cutler A. (2003). Perceptual learning in speech. Cognitive Psychology, 47, 204-238.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (ed.) *Historical Linguistics: Problems and Perspectives* (pp. 237-278). London: Longman.

Oller, K. D. (1973). The effect of position in utterance on speech segment duration in English. Journal of the Acoustical Society of America, 54, 1235–1247.

Peters, B. (2005). Weiterführende untersuchungen zu prosodischen Grenzen in deutscher spontansprache. In K. J. Kohler, F. Kleber, & B. Peters (eds.), *Prosodic structures in German spontaneous speech* (AIPUK 35a) (pp. 203-345). Kiel:IPDS.

Petrone, C., Truckenbrodt, H., Wellmann, C., Holtzgrefe-Lang, J., Wartenburger, I., & Hohle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. Journal of Phonetics, 61, 71-92.

Pierrehumbert, J. B. (1980). The phonology and phonetics of English intonation. MIT Dissertation.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137-157). Amsterdam: John Benjamins.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. Journal of the Acoustical Society of America, 90, 2956-2970.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Roessig, S, & Mücke, D. (2019). Modeling Dimensions of Prosodic Prominence. Frontiers in Commununication, 4, Article 44.

Roessig, S, Mücke, D, & Grice, M. (2019). The dynamics of intonation: Categorical and continuous variation in an attractor-based model. PLoS ONE, 14(5).

Roll, M., & Horne, M. (2011). Interaction of right- and left-edge prosodic boundaries in syntactic parsing. Brain Research, 1402, 93-100.

Roy, J., Cole, J., & Mahrt, R. (2017). Individual differences and patterns of convergence in prosody perception. Laboratory Phonology, 8(1), 22, 1-36.

Sanderman, A. A., & Collier, R. P. G. (1995). Prosodic phrasing at the sentence level. In F. Bell-Berti, & L. J. Raphael (eds.), *Producing speech: Contemporary issues: for Katherine Safford Harris* (pp. 321-332). New York: AIP Press.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. Journal of Phonetics, 52, 183-204.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2016). Individual differences in perceptual adaptability of foreign sound categories. Attention, Perception, & Psychophysics, 78(1), 355-367.

Schertz, J., Kang, Y., & Han, S. (2019). Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast. Laboratory Phonology, 10(1), 13, 1-32. DOI: http://doi.org/10.5334/labphon.67

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. Journal of the Acoustical Society of America, 71, 996-1007.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. Journal of Memory and Language, 57, 24-48.

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. Journal of Psycholinguistic Research, 25, 193-247.

Shattuck-Hufnagel, S., & Turk, A. (1998). The domain of phrase-final lengthening in English. In Proceedings of the 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America, 1235-1236.

Shen, X. (1992). A pilot study on the relation between the temporal and syntactic structures in Mandarin. Journal of the International Phonetic Association, 22, 35-43.

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. Journal of the Acoustical Society of America, 132, EL95-EL101.

Snedeker, J. & Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. Language and Cognitive Processes, 25:7-9, 1234-1264.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. Journal of Memory and Language, 48, 103-130.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. Journal of Psycholinguistic Research, 30(3), 267-295.

Steinhauer, K., Alter, K. & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nature Neuroscience, 2(2), 191-196.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. Brain and Language, 86, 142-164.

Stevens, M., & Harrington, J. (2014). The individual and the actuation of sound change. Loquens, 1(1), e003. DOI: http://dx.doi.org/10.3989/loquens.2014.003

Streeter, L. A. (1978). Acoustic determinant of phrase boundary perception. Journal of the Acoustical Society of America, 64, 1582-1592.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. Journal of the Acoustical Society of America, 101, 514-521.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. Language and Speech, 37(1), 21-43.

Swerts, M., Collier, R., & Terken, J., (1994). Prosodic predictors of discourse finality in spontaneous monologues. Speech Communication, 15, 79-90.

Traxler, M. (2009). A hierarchical linear modeling analysis of working memory and implicit prosody in the resolution of adjunct attachment ambiguity. Journal of Psycholinguistic Research, 38, 491-509.

Truckenbrodt, H., & Féry, C. (2015). Hierarchical organisation and tonal scaling. Phonology, 32, 19-47.

Turk, A. E., & Shattuck-Hufnagel., S. (2007). Multiple targets of phrase-final lengthening in American English words. Journal of Phonetics, 35, 445-472.

Turk, A. (1999). Structural influences on boundary-related lengthening in English. In Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS), San Francisco, 237-240.

van den Berg, R., Gussenhoven, C., & Rietveld, T. (1992). Downstep in Dutch: implications for a model. In G. J. Docherty, & D.R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (p. 334). Cambridge: Cambridge University Press.

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2020). “itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.” R package version 2.4.

Venditti, J. J., Jun, S.-A., & Beckham, M. E. (1996). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In J. L. Morgan & K. Demuth (eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 287-311), Mahwah: Lawrence Earlbaum Associates.

Weber, A., Grice, M., & Grocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: a study of anticipatory eye movements. Cognition, 99 (2), B63-B72.

Wellmann, C., Holtzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. Frontiers in psychology, 3, Article 580.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wightman C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. Journal of the Acoustical Society of America, 91, 1707–1717.

Wood, S. N. (2006). 'Low-rank scale-invariant tensor product smooths for generalized additive mixed models. Biometrics, 62(4), 1025-1036. DOI: https://doi.org/10.1111/j.1541-0420.2006.00574.x

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B), 73(1), 3-36.

Yang, X., Shen, X., Li, W., & Yang, Y. (2014) How listeners weight acoustic cues to intonational phrase boundaries. PLoS ONE, 9(7), e102166. DOI: 10.1371/journal.pone.0102166

Zhang, X. (2012). A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese. University of Michigan dissertation.

Zvonik, E., & Cummins, F. (2002). Pause duration and variability in read texts. In Proceedings of Int. Conf. Spoken Language Processing (ICSLP-2002), 1109-1112.

Zvonik, E., & Cummins, F. (2003). The effect of surrounding phrase lengths on pause duration. In Proceedings of Eurospeech-2003, 777-780.