Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Prosodic boundary strength: An articulatory and perceptual study

Jelena Krivokapić [a,*], Dani Byrd [b]

[a] Department of Linguistics, Yale University, 370 Temple St, Room 204, New Haven, CT 06520-8366, USA
[b] Department of Linguistics, University of Southern California, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, USA

ARTICLE INFO

ABSTRACT

This study examines the production and perception of Intonational Phrase (IP) boundaries. In particular, it investigates (1) whether the articulatory events that occur at IP boundaries can exhibit temporal distinctions that would indicate a difference in degree of disjuncture, and (2) to what extent listeners are sensitive to the effects of such differences among IP boundaries. Two experiments investigate these questions. An articulatory kinematic experiment examines the effects of structural differences between IP boundaries on the production of those boundaries. In a perception experiment listeners then evaluate the strength of the junctures occurring in the utterances produced in the production study. The results of the studies provide support for the existence of prosodic strength differences among IP boundaries and also demonstrate a close link between the production and perception of prosodic boundaries. The results are discussed in the context of possible linguistic structural explanations, with implications for cognitive accounts for the creation, implementation, and processing of prosody.

1. Introduction

Within theories of prosodic structure there are disagreements on the number and definition of prosodic constituents, but in general all theories of English assume at least a major (large) and a minor (small) prosodic category above the level of word. These have been codified as Intonational Phrase (IP) and the Phonological or Inter-mediate Phrase (ip), respectively (as suggested in Beckman & Pierrehumbert, 1986; for an overview see Shattuck-Hufnagel & Turk, 1996). A number of studies have investigated temporal properties at the edges of these phrases. Acoustic studies have shown that at boundaries segments increase in duration (e.g., Gaitenby, 1965; Klatt, 1975; Oller, 1973; Turk & Shattuck-Hufnagel, 2007; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Articulatory studies have shown that speech gestures become temporally longer in the vicinity of boundaries, and that articulatory lengthening increases cumulatively, distinguishing several levels of phrasal lengthening (e.g., Byrd & Saltzman, 1998; Byrd, Kaun, Narayanan, & Saltzman, 2000; Byrd, 2000; Cho, 2006; Edwards, Beckman, & Fletcher, 1991; Tabain, 2003).

While these properties of prosodic boundaries are well-established, it has long been acknowledged that there are differences among boundaries that can be considered to be of the same category. As Ladd (1996:239) points out, a small number of

categories does not correspond to the actual productions of prosodic categories and does not capture the distinctions listeners are able to make between boundaries of different strength. Speakers seem to be able to produce and listeners to perceive much finer grained structures. Little experimental evidence exists for these observations however, and it remains an open question (1) whether among phrase boundaries with like intonational patterns, which would be all considered a large or IP boundary, there are quantitative distinctions in the articulation that indicate a difference in boundary strength, i.e., degree of disjuncture and (2) whether, if such differences exist in the production, listeners are sensitive to them. A related question is: If there are systematic differences in boundary strength, how are such differences represented linguistically? There are two possible answers: (1) to assume different prosodic categories for each boundary strength, and (2) to assume prosodic recursion. We for now assume the framework of Beckman and Pierrehumbert (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986) and will return to these questions in the discussion section.

The empirical basis for the discussion so far is a few acoustic studies and perceptual studies that indirectly address boundary strength differences. Wightman et al. (1992) examine the seven boundary types suggested in Price, Ostendorf, Shattuck-Hufnagel, and Fong (1991). Based on final lengthening of the vowel in phrase final syllables (across different syllable structures), Wightman et al. (1992) find evidence for four distinct prosodic categories. However, they point out that a larger number of prosodic boundaries might be distinguished if cues other than final lengthening are taken into account, for example pauses or pitch. The lengthening of the phrase final coda consonant, shown

in Fig. 4a in Wightman et al. (1992), also seems to point to five boundary strength distinctions. These data indicate that boundaries classified in later prosodic hierarchies as large/IP boundaries might differ in strength. Indeed, Ladd (1988) reports phonetic evidence (declination and boundary duration, where boundary duration was measured as the duration from the onset of the last stressed syllable to the onset of postboundary phonation[1]) showing that IP boundaries can differ in strength. Wagner (2005) reports an acoustic study examining boundary strength in coordinative structures, and although the used measures are not sufficiently fine-grained to examine boundary properties in detail (the duration of whole constituents was measured), the results again suggest boundaries of different strength. Frota (2000) finds experimental evidence (based on acoustic duration, F0, and sandhi phenomena) of IP strength differences in European Portuguese.

On the perceptual side, no study has evaluated directly the question of whether large/IP boundaries allow for finer grained distinctions among boundaries that all display the gross characteristics of final lengthening and boundary tone. An indication that there might be differences within a category comes from a study by Wagner and Crivellaro (2010), in which it was shown, using synthesized speech in a syntactic disambiguation task, that there is a correlation between gradient boundary strength and listeners' choice of syntactic structure. Although in this study it was not clear to what extent the differences in boundary strength were within the same category and to what extent between different categories, the study did show that as boundary strength gradually increases, the likelihood of listeners perceiving that disjuncture as a syntactic boundary increases as well. Thus the study indicates that differences within a category might be relevant in boundary perception. De Pijper and Sanderman (1994) examine a large database of boundaries in an experiment testing the relation between phonetic cues and perceived boundary strength, and they notice that the perceived strength of boundaries does not seem to form a small number of clusters. As they point out, these results suggest that boundaries of the same prosodic category can have different strengths. Similarly, Swerts (1997) in a study of discourse boundaries finds that listeners distinguish six degrees of boundary strength as a function of pause duration. Finally, Krivokapić and Ananthakrishnan (2007) examine the perception of a wide range of prosodic boundaries and find that listeners perceive five distinct categories, again more than expected given the none/word vs. minor/small vs. major/large gross categories.

The above studies all indicate, from the point of view of both production and perception, that prosodic categories of like general type can differ in strength, both as created by the speaker and as interpreted by the listener. However, the empirical evidence is scant, and comes, in production studies, mostly from F0 data and from acoustic duration measures of large temporal intervals; these measures might not be fine-grained enough to evaluate the question of whether like-type prosodic categories can be implemented in systematically different ways. Only indirect evidence exists from the perception of prosodic structure. We present two experiments explicitly testing this question in English. Structures of the kind Ladd (1986, 1996) proposed and investigated in the tonal and acoustic temporal domain (Ladd, 1988) will be examined in the temporal domain in articulation and perception.

By evaluating both production and perception we aim to investigate whether the distinctions that speakers produce are also perceived by the listeners. Examining articulation allows us a more direct view into what are likely to be small differences in boundary strength. If differences exist, evidence from perception will be crucial in establishing their structural relevance, as opposed to these differences being random variations in production (such as small variation in VOT might be, for example).

2. Experiments

The first part of this study is an articulatory magnetometer study investigating the production of boundaries in four sentences in which the large phrase boundaries result from like syntactic structures. In the second part of the study listeners evaluate the perception of these same boundaries. In the production part of the study, articulatory kinematic data were collected simultaneously with audio recordings, and these audio recordings were used as stimuli in the perception part of the study. The goal of the experiments is to examine whether IP boundaries in English show systematic strength differences that are distinguishable—in a parallel way—in both production and perception.

2.1. Methods: production

2.1.1. Stimuli and subjects

Four sentences examining the boundary strength differences in two different phonetic contexts were constructed, shown in Table 1. To help subjects put the sentences in a sensible context, the sentences were preceded by a question, which the subjects did not read aloud.

We examine the consonantal constrictions at the boundaries marked with the pound sign, named boundary A and boundary B, and in sentences 3 and 4 the vocalic constrictions as well. The consonants in the first sentence type (sentences 1 and 2) are alveolar stops, and in the second sentence type the first consonant is an alveolar fricative, the second an alveolar stop (the different consonantal constrictions were included so as to make sure that the results are not specific to one constriction). The sentences are expected to be produced as three ToBI IPs, shown in Table 2 (the prosody verification procedure and results will be discussed in Section 2.1.4).

Semantically and syntactically, the sentence pairs are similar in that they have a temporal interval (1 and 2) and spatial interval (in 3 and 4) as an argument (Roumyana Pancheva, p.c., see also the discussion on Extent arguments in Dowty, 1991). The two sentences in each pair do not differ truth-conditionally, and any potential difference in meaning between them derives from a difference in information structure. For example, if the *to*-phrase is first, as in sentence 1, the relevance of the end of the work shift is emphasized; whereas if the *from*-phrase is first, the order of the temporal interval from the starting point to the end point is preserved, as in sentence 2 (Laurence Horn, p.c.). A similar difference in meaning exists in sentences 3 and 4.

We will refer to the boundary between the first two IPs as boundary A and between the second two IPs as boundary B. Crucially both boundaries are preceded and followed by boundaries that are marked by a boundary tone and final lengthening (i.e., by boundaries that exhibit the characteristics of an IP boundary). The test sentences could give rise to three possible situations: boundary A and B could be non-distinct (i.e., of equal strength or disjuncture), boundary A could be systematically stronger than boundary B, or boundary B could be systematically stronger than boundary A. To put this another way, the break, as examined in the articulation, might be bigger between the first

---

[1] Ferreira (1993) shows that pause duration depends in part on final lengthening, in that more lengthening leads to shorter pauses, and that to what extent the boundary will be realized as lengthening or as a pause depends on the phrase final segments. The implications of this finding are, as Ferreira argues, that both lengthening and pausing are instantiations of prosodic boundaries. See also Byrd and Saltzman (2003) for a prosodic boundary model that can integrate these findings.

Table 1
Stimuli for the production study. The context sentence was not read aloud by the subjects. The boundary of interest is marked by the pound sign (there was no pound sign in the experimental stimuli presented to the subjects).

	Stimuli	Condition	Segmental context
Context:	*What is the workers' shift?*		
Subject:	1. They usually do: # to 2:00, from 8:00	Boundary A	/**d** u **t** ə t u/
Subject:	2. They usually do: from 2:00, # to 2:00	Boundary B	/**t** u **t** ə t u/
Context:	*What range on the map grid do they have?*		
Subject:	3. They usually see: # to 2, from C	Boundary A	/**s i t ə** t u/
Subject:	4. They usually see: from C, # to 2	Boundary B	/**s i t ə** t u/

Table 2
Prosodic phrasing.

1. [$_{IP1}$ They usually do]: [$_{IP2}$ to 2:00], [$_{IP3}$ from 8:00]
2. [$_{IP1}$ They usually do]: [$_{IP2}$ from 2:00], [$_{IP3}$ to 2:00]
3. [$_{IP1}$ They usually see]: [$_{IP2}$ to 2], [$_{IP3}$ from C]
4. [$_{IP1}$ They usually see]: [$_{IP2}$ from C], [$_{IP3}$ to 2]

two IPs or between the second two IPs; or there might be no difference. Given the semantic and syntactic structure of these sentences, there is no strong a priori reason that only one of these patterns could arise.[2]

The literature strongly suggests (e.g., Byrd & Saltzman, 1998; Byrd et al., 2000; Byrd, Krivokapić, & Lee, 2006; Cho, 2006; Edwards et al., 1991; Tabain, 2003) that articulatory constriction duration is strongly related to boundary strength, in that articulatory lengthening increases with boundary strength. We will use this articulatory measure as a critical probe for differences. Crucially, in either of the two readings in which the IP boundaries differ in strength, boundary-adjacent constriction duration can track that difference. The null hypothesis is that the two examined boundaries will be of equal strength. However, based on the small set of acoustic studies (Frota, 2000; Ladd, 1988; Wagner, 2005; Wightman et al., 1992), we expect that the two boundaries may differ in strength.

In addition to the test sentences, two more sentences were added as an intended control condition. They place the relevant segment in phrase medial context, and we expected, based on previous work (e.g., Byrd & Saltzman, 1998; Byrd et al., 2006; Cho, 2006; Edwards et al., 1991; Fougeron & Keating, 1997; Tabain, 2003) that constrictions in the test conditions would be longer than the constrictions in the no-boundary control condition. However, as it turned out, the two sentences were all produced with a sentence-medial boundary (ip or IP).[3] Since in order to be control sentences a no-boundary reading would be required, these are not used in further analysis,

nor in the perception study. They will not be discussed further, except as necessary for understanding the experiment procedure.

2.1.2. Subjects and procedures

Four native speakers of American English with no known speech or hearing disorders participated in the study. Data for a fifth speaker were originally collected, but are not used since the labelers disagreed on the prosodic labeling for this subject (see the description on labeling in Section 2.1.4). Subjects were paid for their participation and were naïve as to the purpose of the experiment. One subject's articulatory data (subject J) were unusable due to excessive magnetometer tracking error. (The data did not show regular tongue movement patterns. We suspect that the subject leaned forward during the experiment, thereby moving out of the optimal measurement range in which the tongue movement could reliably be tracked.) However, the speaker's audio data were prosodically labeled (Section 2.1.4) and acoustically analyzed (Section 2.4), and are used in the perception study. Subjects will be referred to as subject D, K, J, M.

Each sentence was repeated 10 times for the first subject (subject J), and 13 times for the other subjects, for a total of 60 tokens for subject J (3 prosodic conditions [two test and one control condition] × 2 segmental contexts × 10 repetitions) and 78 tokens per subject for the 3 other subjects (3 prosodic conditions × 2 segmental contexts × 13 repetitions). The sentence repetitions were randomized in blocks. They were blocked by sentence (sentence 1, sentence 2, etc.) and by condition (boundary A, boundary B, no-boundary). In order to decrease variability in the production of the sentences, both no-boundary sentences were read at the end of the experiment, not intermingled with the prosodically complex sentences, which blocked could appear in either order, A then B or B then A. Within each prosodic condition, the repetitions for the two different sentences were blocked, but the order of the two sentence types was random, thus for example for the condition boundary A, subjects would read either all repetitions of sentence 1 followed by all repetitions of sentence 3, or the other way around.

Subjects were given a context for the sentences (two examples are shown in Appendix I). For the first set of sentences (the 'do to' sentences), the context sentence was: "What is the workers' shift?". This sentence was shown on a sheet of paper with a picture of a clock on it and with the stimulus sentence as the answer. For the second set of sentences (the 'see to' sentences) the context sentence was: "What range on the map grid do they have?". This sentence was presented with a picture of a grid on the paper and with the stimulus sentence as the answer. During the experiment, the appropriate sheet with the context question and picture preceded each of the six blocks of sentences.

Since the utility of the data depends on speakers generating IP boundaries for all three target junctures in the sentences, the instructions and presentation were oriented to yield such a reading. To help elicit these large phrase boundaries, the three

---

[2] It is well known that prosodic structure is to a large extent determined by syntactic structure, but it is also known that it varies by speaker (see e.g., Yoon (2007:29ff) for variability in production of prosody by radio speakers reading the same text; Campione and Veronis (2002) and Cole, Mo, and Baeck (2010) for variability in a corpus study of spontaneous speech. As we do not see any truth-conditional difference in meaning between the two possible phrasings (i.e., boundary A being stronger than boundary B, or boundary B being stronger than boundary A), the choice between them is likely to be determined by speaker specific preferences, for example by their rhythmic preferences, or by their preference as to which constituent they want to emphasize by the grouping (in the case of these sentences, either emphasizing the spatial/temporal interval argument, or emphasizing the last constituent). A further possibility is that the phrasing reflects how incrementally individual speakers process sentences. (See Swets, Desmet, Hambrick, and Ferreira (2007) on the relationship between prosodic phrasing and individual differences in processing. See also Nespor and Vogel (1986) on prosodic restructuring.)

[3] The two sentences were "2 to 2" and "C to 2", to be used as control for sentences 1, 2, and 3, 4 respectively. The same context questions were given as for sentences 1, 2 and 3, 4.

phrases were visually distinguished by using capital, small caps and regular fonts for the first, second, and third IP respectively. The control sentence was printed in regular font.

The instructions for reading the test sentences were: "Please read the question silently, and the answers aloud. Please read carefully, paying attention to punctuation." For the control sentences the instructions were: "Please read the question silently, and the answers aloud. Please read carefully, and do *not* pause within the sentence." The instruction sheet for two sentences (one for each sentence type) is given in Appendix I.

In order to further ensure the production of three IPs, subjects were given sentences similar to the experimental sentences in a practice session and asked to read them aloud (the materials for the practice sentences are given in Appendix II). The instructions for the practice session were the same as for the recording session.

During the experiment, in case the subjects did not read the test sentences as three IPs, the experimenter asked them to read the sentence again and pay attention to the punctuation. In case subjects paused in the control sentence, the experimenter asked them not to. Neither of these situations arose very often.

2.1.3. Data collection

Data were collected using the 3-dimensional Carstens Articulograph system (AG500) with seven sensors: four on the tongue (tongue tip, tongue body, tongue dorsum, and tongue rear) tracking articulatory tongue movements, and three reference sensors (one on the nose and two behind the ears) for head movement correction. The articulatory data were sampled at 200 Hz and the acoustic data at 16 kHz and pre-processed using CalcPos (the description of the program is available at: http://www.ag500.de/calcpos/CalcPos_2.pdf). The data were corrected for head movement and rotated to the occlusal plane, so that the *x–y* plane of the data coordinate system aligned with subject's occlusal plane. In order to derive the vertical velocity signal, the *z* (vertical) signal component was differentiated for the four tongue sensors. The signals were smoothed before and after differentiation with a 9th order Butterworth filter of cutoff frequency 15 Hz in order to remove the high frequency noises of the sensor trajectories. The horizontal ($\times$) velocity signal was also derived similarly and the tangential velocity in the midsagittal plane was computed by combining the horizontal and the vertical velocity components.

2.1.4. Data analysis

Two trained prosodic labelers, one the first author and one a labeler naïve as to the purposes of the study, transcribed the audio recordings following the ToBI transcription guidelines (Beckman & Ayers Elam, 1997). Since the study investigates differences in boundary strength within IP boundaries, it is important to establish that the boundaries examined are indeed IPs, so the prosodic boundary markings distinguishing ip and IP boundaries are crucial. The criteria for an ip boundary are phrase final lengthening (indicated by a break index 3 in the ToBI labeling guide) and a phrase accent, and for an IP boundary phrase final lengthening (indicated by a break index 4 in the ToBI labeling guide), a phrase accent, a boundary tone, and optionally the occurrence of a pause. Although both ip and IP are marked by final lengthening, IP boundaries lengthen more. Boundary tones can be difficult to discern in certain cases (most notably in the situation when both the phrase accent and the boundary tone are low, so in L–L%). In the data in this study for subjects J and K the tone sequences were nearly always L–H% at the relevant boundaries. Subject M had the tone sequence L–L% at the first relevant boundary in sentences 1, 2, and 3 (after *They usually do/They usually see*), and the tone sequence L–H% at all the other relevant boundaries. Subject D had L–H% and H–L% at the relevant boundaries, and some L–L% sequences. Only data where both labelers agreed on the boundary type (indicated by the tonal properties and by a break index 4) were used in the study. In total, 21 out of 234 sentences (3 subjects $\times$ 6 sentences $\times$ 13 repetitions) for the three subjects used in the production study were discarded due to labeler disagreement, and none for subject J. All other test sentences were produced as three IPs, thus allowing us to test whether the two IP boundaries (marked as A and B above) differ in strength. As mentioned before, the control sentences were produced with a boundary and excluded from further analysis. Thus, there were 135 sentences analyzed in total (3 subjects $\times$ 4 sentences $\times$ 13 repetitions, minus the 21 discarded sentences).

The articulatory signals were analyzed using MView software (M. Tiede under development). The vertical velocity trajectories of the articulator position trajectories were used to identify important kinematic landmarks in the consonant constriction formation (Fig. 1 shows the labeling for consonants and vowels). For each consonant constriction, data were available for the tongue tip and the tongue blade movement. For two subjects (D and M) tongue tip data were used; for one subject (subject K) tongue blade data
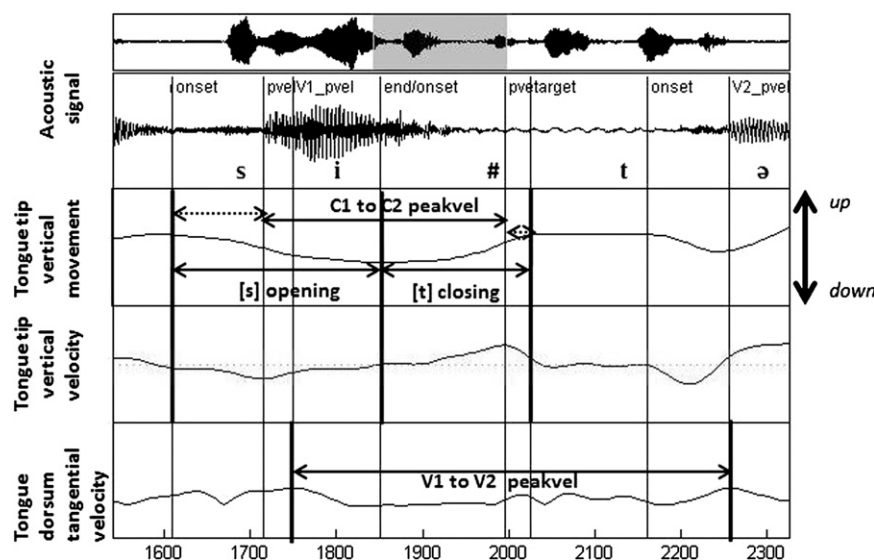
Fig. 1. An example of the identified landmark events and the derived variables. The vertical lines show the identified landmarks (as described in the text). The horizontal full lines show the identified variables, and the dotted lines the acceleration durations (C1 opening movement acceleration duration and C2 closing movement deceleration duration).

were used since the tongue tip data contained many tracking errors. Six time points were identified from the velocity signal for the consonant constriction forming movements: the onset of the preboundary opening movement, the end of the preboundary opening movement (which is also by definition the onset of the postboundary closing movement), the target of the postboundary constriction closing movement, the onset of the postboundary constriction opening movement (the target and the onset of the opening movement can be identical; when they are not identical, the closure is plateau shaped), the peak velocity time points for the preboundary opening and for the postboundary closing movement. From these time points five dependent variables were derived that reflect different aspects of the duration of the consonant gestures (C1 refers to the preboundary consonant, C2 to the postboundary consonant):

- *C1 opening movement acceleration duration*: the time from the onset of the preboundary opening movement to the peak velocity of that movement. (Acceleration/deceleration duration is an indicator of gestural stiffness.)
- *C1 opening movement duration*: the time from onset to the end of the opening movement
- *C2 closing movement deceleration duration*: the time from peak velocity to the target of the closing movement. (The deceleration duration parallels acceleration duration but was chosen for C2 because it was a more stable measure.)
- *C2 closing movement duration*: the time from closing movement onset to the target of the closing movement. We also examined the time from closing movement onset to the midpoint between target of closing movement and onset of opening movement, but since the results are substantively identical for both measures, we report only results for the first measure.
- *C1-to-C2 peakvel*: from the C1 opening movement peak velocity to the C2 closing movement peak velocity. In the data, these seemed to be stable points across repetitions, and given that the variable spans the boundary (more so than other measures) we take this variable to be a measure of the boundary interval duration.

For the vowels (in sentences 3 and 4), the movement of the tongue dorsum was tracked. The tangential velocity trajectories were used to identify the peak velocity for the closing movement for the preboundary vowel and for the postboundary vowel. In the case of multiple peaks, the highest peak was chosen. From this a further dependent variable was derived (V1 refers to the preboundary and V2 to the postboundary vowel):

- *V1-to-V2 peakvel*: from the V1 closing movement peak velocity, associated with the [i], to the V2 maximal retraction movement peak velocity, associated with the [ə]. As with the consonants, this can be taken as a measure of boundary duration.

2.1.5. Statistical analysis

A set of within-subject two factor ANOVAs was conducted using the statistical software package Statview (produced by the SAS Institute), testing the effect of boundary (with two levels: 'boundary A' and 'boundary B') and consonant (with two levels: 'alveolar stop [t]/[d]' and '[s]') on the dependent variables. The dependent variables were: C1 opening movement acceleration duration, C1 opening movement duration, C2 closing movement deceleration duration, C2 closing movement duration, and C1-to-C2 peakvel. For the V1-to-V2 peakvel interval a one factor ANOVA testing the effect of boundary (with the two levels: 'boundary A' and 'boundary B') was conducted. Criterial significance was set at $p < .05$. All and only statistically significant results are reported.

2.2. Results: production

Table 3 shows the results for all three speakers, for all variables in the production experiment. Overall, the results show that each speaker does produce IP boundaries of different strengths.

Subjects K and M have a stronger boundary A than boundary B. Subject K's differences in boundary strength are reflected in the significantly longer durations for boundary A than for boundary B for the following measures: C1 opening movement, the vowel boundary duration measure (V1-to-V2 peakvel), and the consonant boundary duration measure (C1-to-C2 peakvel). There are further significant effects of the consonant such that the preboundary opening movement, preboundary acceleration duration, and C1-to-C2 peakvel are longer when C1 is [t]/[d] than when C1 is [s]. Finally, there is a nearly significant boundary effect for the C1 acceleration duration, such that boundary A is weaker than boundary B. This effect is driven by [t]/[d], as the effect of consonant [s] is in the opposite direction as shown by the interaction of consonant and boundary effects for the C1 acceleration duration. Subject M also has a stronger boundary A than boundary B, as shown in the significantly longer vowel boundary duration measure (V1-to-V2 peakvel) and in the nearly significantly longer consonant boundary duration measure (C1-to-C2 peakvel) for boundary A than for boundary B. However, boundary B is stronger than boundary A in the preboundary acceleration interval. There is also an interaction of consonant and boundary effects, such that the preboundary opening movement is longer at boundary A than boundary B when C1 is [s], and the opposite is the case when C1 is [t]/[d]. In addition to these effects, there is an effect of the consonant such that the C1 opening movement and the C1 acceleration interval are longer for [t/d] than for [s].

Subject D shows the opposite direction of boundary strength, having a stronger boundary B than boundary A, as shown in the C1 opening movement and the consonant boundary duration measure (C1-to-C2 peakvel). Both are significantly longer for boundary B than for boundary A. For the C1-to-C2 peakvel, the boundary only has an effect when C1 is [t]/[d] (as shown by the interaction between consonant and boundary effect for C1-to-C2 peakvel). In addition, there is an effect of consonant, such that C1-to-C2 peakvel is longer when C1 is [t]/[d] than when it is [s]. There is an interaction between consonant and boundary effects for the C1 opening movement, such that the boundary effect is stronger for consonants [t]/[d] than for [s]. A consonant and boundary interaction effect is found for the C2 closing movement deceleration duration, such that the boundary strength trends go in opposite directions for the two consonants.

Summarizing, we see overall an effect of boundary, speaker, and consonant. The main effect of the consonant is that [t/d] is longer than [s], which is not surprising given that it has previously been found that movement amplitude differs in these consonants, such that [d] and [t] show more displacement than [s] (Fuchs, Perrier, Geng & Mooshammer, 2006; Mooshammer, Hoole, & Geumann, 2006). Since temporally longer movement typically leads to larger displacement, the observed effect of consonant can be explained by different kinematic properties of these consonants. Overall the results indicate that for each of the three speakers, the two boundaries are of different strength. Boundary A was produced as stronger for speakers K and M, and boundary B was produced as stronger for speaker D. Thus, speakers do differ in which of the three IPs group or cohere more strongly together, that is, in whether they choose to use a phrasing with A stronger than B or the reverse. *Crucially*, either pattern contradicts the null hypothesis that the two IP boundaries emerge alike in boundary strength. We turn now to the perception experiment, in which we analyze how the acoustic data produced by these articulatory movements are perceived by listeners.

Table 3

Means (SD) for all variables in the production experiment. Measurements given in ms. Significance is marked with * and with bolded fonts, non-significant results are given in regular fonts.

	C1 opening movement duration	C1 opening movement acceleration duration	C2 closing movement deceleration duration	C2 closing movement duration	C1-to-C2 peakvel	V1-to-V2 peakvel
Subject K						
Effect of boundary	**$F_{(1,30)}$=5.926, p=.0211*** **Boundary A=334 (233)** **Boundary B=254 (91)**	$F_{(1,30)}$=4.169, p=.0501 Boundary A=50.3 (17.5) Boundary B=80 (33.4)	$F_{(1,30)}$=.821, p=.3721 Boundary A=91.1 (44.9) Boundary B=77.8 (27.8)	$F_{(1,30)}$=2.850, p=.1018 Boundary A=238.9 (62.5) Boundary B=205.9 (43.4)	**$F_{(1,30)}$=8.227, p=.0072*** **Boundary A=420 (223)** **Boundary B=299 (75)**	**$F_{(1,17)}$=50.947, p < .0001*** **Boundary A=646.3 (100.7)** **Boundary B=379.4 (51.6)**
Effect of consonant	**$F_{(1,30)}$=13.588, p=.0009*** **[t]/[d]=366 (185)** **[s]=197 (126)**	**$F_{(1,30)}$=17.756, p=.002*** **[t]/[d]=77.2 (28.8)** **[s]=45.7 (20.3)**	$F_{(1,30)}$=.598, p=.4452 [t]/[d]=77.4 (39.1) [s]=75.4 (34.9)	$F_{(1,30)}$=2.072, p=.1604 [t]/[d]=206.3 (35.5) [s]=247.7 (71)	**$F_{(1,30)}$=8.319, p=.0072*** **[t]/[d]=411.2 (195.9)** **[s]=293.6 (128.2)**	N/A
Consonant and boundary interaction effect	$F_{(1,30)}$=.181, p=.6740 Boundary A [t]/[d]=464 (272) [s]=229 (132) Boundary B [t]/[d]=301 (29) [s]=115 (63)	**$F_{(1,30)}$=17.032, p=.0003*** **Boundary A** **[t]/[d]=50.6 (12.1)** **[s]=50 (21.6)** **Boundary B** **[t]/[d]=95 (22)** **[s]=35 (13.5)**	$F_{(1,30)}$=1.079, p=.3072 Boundary A [t]/[d]=85.4 (44.2) [s]=56.8 (24.4) Boundary B [t]/[d]=68.5 (32.7) [s]=92.5 (35.1)	$F_{(1,30)}$=3.795, p=.0608 Boundary A [t]/[d]=203.5 (24) [s]=267.3 (70.1) Boundary B [t]/[d]=208.3 (42.5) [s]=198.7 (52)	$F_{(1,30)}$=.446, p=.5093 Boundary A [t]/[d]=532 (269) [s]=329 (132) Boundary B [t]/[d]=330 (48) [s]=204 (60)	
Subject D						
Effect of boundary	**$F_{(1,40)}$=9.303, p=.0040 *** **Boundary A=389 (101)** **Boundary B= 473 (95)**	$F_{(1,40)}$=.785, p=.3810 Boundary A=64.5 (29.7) Boundary B=75 (39.9)	$F_{(1,40)}$=.789, p=.3798 Boundary A=71.136 (37.8) Boundary B=81.591 (35.4)	$F_{(1,40)}$=1.593, p=.2142 Boundary A=219.9 (77.7) Boundary B=194.2 (62.1)	**$F_{(1,40)}$=4.244, p=.0459*** **Boundary A=439 (98)** **Boundary B=485 (85)**	$F_{(1,24)}$=.712, p=.4072 Boundary A=637.3 (151) Boundary B=599.2 (60.3)
Effect of consonant	$F_{(1,40)}$=.215, p=.6457 [t]/[d]=435 (145) [s]=426 (51)	$F_{(1,40)}$=3.613, p=.0645 [t]/[d]=59.3 (25.9) [s]=79.4 (40.1)	$F_{(1,40)}$=.048, p=.8271 [t]/[d]=77.4 (39.2) [s]=75.4 (34.9)	$F_{(1,40)}$=.516, p=.4768 [t]/[d]=215.3 (77.7) [s]=199.6 (54.6)	**$F_{(1,40)}$=4.350, p=.0434*** **[t]/[d]=488.3 (112.5)** **[s]=438.3 (66.6)**	N/A
Consonant and boundary interaction effect	**$F_{(1,40)}$=4.461, p=.0410*** **Boundary A** **[t]/[d]=365 (129)** **[s]=412 (59)** **Boundary B** **[t]/[d]=513 (125)** **[s]=439 (41)**	$F_{(1,40)}$=.716, p=.4025 Boundary A [t]/[d]=59.1 (8.6) [s]=70 (41.3) Boundary B [t]/[d]=59.5 (37.5) [s]=87.9 (38.6)	**$F_{(1,40)}$=6.230, p=.0168*** **Boundary A** **[t]/[d]=85.4 (44.2),** **[s]=56.8 (24.4)** **Boundary B** **[t]/[d]=68.5 (32.7)** **[s]=92.5 (35.1)**	$F_{(1,40)}$=.237, p=.6289 Boundary A [t]/[d]=232 (85) [s]=207.7 (50.3) Boundary B [t]/[d]=196.7 [s]=192.1 (59.5)	**$F_{(1,40)}$=6.618, p=.0139*** **Boundary A** **[t]/[d]=433 (114)** **[s]=445 (84)** **Boundary B** **[t]/[d]=549 (76)** **[s]=432 (47)**	N/A
Subject M						
Effect of boundary	$F_{(1,40)}$=.007, p=.934 Boundary A=278 (82) Boundary B=279 (123)	**$F_{(1,40)}$=8.870, p=.0049*** **Boundary A=136.6 (73.2)** **Boundary B=174.1 (106.3)**	$F_{(1,40)}$=.915, p=.3447 Boundary A=83.2 (41) Boundary B=95.7 (42.7)	$F_{(1,40)}$=1.426, p=.2394 Boundary A=240.8 (47.9) Boundary B=222.6 (45.6)	$F_{(1,40)}$=3.995, p=.0525 Boundary A=260.9 (72.3) Boundary B=218.4 (75.9)	**$F_{(1,24)}$=15.692, p=.0006*** **Boundary A=615.4 (75.6)** **Boundary B=488.8 (86.9)**
Effect of consonant	**$F_{(1,40)}$=20.568, p < .0001*** **[t]/[d]=331 (72)** **[s]=220 (103)**	**$F_{(1,40)}$=108.004, p < .0001*** **[t]/[d]=226.7 (59.8)** **[s]=77.1 (45.2)**	$F_{(1,40)}$=.097, p=.7569 [t]/[d]=91.1 (32.8) [s]=87.6 (50.7)	$F_{(1,40)}$=.826, p=.3690 [t]/[d]=238.1 (49.7) [s]=224.8 (44.2)	$F_{(1,40)}$=1.213, p=.2773 [t]/[d]=228.9 (73.8) [s]=251.4 (79.1)	N/A
Consonant and boundary interaction effect	**$F_{(1,40)}$=9.581, p=.0036*** **Boundary A** **[t]/[d]=294 (50)** **[s]=259 (108)** **Boundary B** **[t]/[d]=372 (71)** **[s]=185 (88)**	$F_{(1,40)}$=2.113, p=.1538 Boundary A [t]/[d]=195.8 (30.2) [s]=65.5 (33.1) Boundary B [t]/[d]=260.5 (66.9) [s]=87.7 (53.4)	$F_{(1,40)}$=.412, p=.5246 Boundary A [t]/[d]=81.2 (31.7) [s]=85.5 (51.7) Boundary B [t]/[d]=101.8 (31.9) [s]=89.5 (52.1)	$F_{(1,40)}$=2.935, p=.0944 Boundary A [t]/[d]=257.3 (43.7) [s]=221 (47.2) Boundary B [t]/[d]=217 (49.1) [s]=228.2 (43.3)	$F_{(1,40)}$=.999, p=.3235 Boundary A [t]/[d]=240 (44) [s]=286 (92) Boundary B [t]/[d]=217 (98) [s]=219 (50)	N/A

2.3. Methods: perception

2.3.1. Stimuli and subjects

The perception experiment examines whether listeners perceive strength differences among IP/large boundaries. The fact that the speakers produced different prosodic phrasings in the production study allows us to specifically examine whether in the perception listeners track the production of the speakers, rather than, for example, using their own 'internal' prosodic structure for these sentences or some sort of default prosodic phrasing. The goal of the perception study is thus to examine how the individual differences in the prosodic phrasing of the speakers are reflected in the perception of the listeners.

Audio recordings collected for the four test sentences from the production study were used as stimuli in the perception part of the study. Data from four speakers were used, namely subjects D, K, M, and subject J (whose articulatory data were not tracked but whose acoustic data were measured and prosodically labeled). Only sentences that were used in the production study were evaluated, and one sentence token was excluded due to experimental error. In total, from the 196 sentences for 4 subjects pooled (4 sentences × 10 repetitions for subject J, 4 sentences × 13 repetitions for subjects D, K, M), 22 sentences were excluded (13 for subject K, 3 for subject M, and 6 for subject D), so there were 174 sentences for the perception study.

2.3.2. Data collection

Listeners evaluated boundary strength by using a computer version of the Visual Analogue Scale (VAS, see for example Rietveld & Chen, 2006; Wewers & Lowe, 1990). In VAS studies, subjects are presented on a screen or paper with a line, with the ends of the line marked for the phenomena measured (e.g., 'no pain' and 'highest pain'). The subjects mark their estimate of the strength of the stimuli (e.g., strength of pain) on the line. The results are evaluated by measuring the distance of the mark from one end of the scale. For this study, a computer version of the VAS, as implemented by Granqvist (1996), was used. The instructions to the listeners were as follows:

"You will hear a number of sentences. In each sentence, one of the following phrases will appear:

do to
two to
see to
C to

Please judge how strongly connected the two words are. You can listen to the sentences two times. When you have decided, give your answer by clicking on the bar on the screen."

The ends of the scale were marked as 'weakest connection' (indicating the strongest boundary or disjuncture) and 'strongest connection' (indicating no prosodic boundary due to two strongly connected words). This naming convention for the ends of the scale was used so that listeners did not need to be trained on the notion of 'boundary' and could simply report how strongly they felt the two words to be connected.

Listeners were given a practice trial, with the purpose of exemplifying the use of 'strong connection' and 'weak connection' and making them familiar with using the scale. During the practice trial listeners heard the four stimulus sentences of the experiment, each spoken once with a very strong boundary and once with no boundary at all. The eight sample sentences were spoken by a speaker different from the speakers in the production part of the study.

In the experiment, each listener listened to all 174 sentences, blocked by speaker. The sentences were randomized, i.e., not blocked by sentence or by condition. There were four parts of the experiment (each part consisting of sentences produced by one speaker), with a one-minute break between each part. 25 listeners were asked to rate the strength of boundaries. They were native speakers of American English with no known speech or hearing impairments.

2.3.3. Data analysis

For each listener's response, the software returns a numerical value on a scale of 0–1000. For the 'weakest connection' (strongest boundary) 0 is returned, and 1000 is returned for 'strongest connection' (weakest boundary). Since listeners might differentially use different ranges of the scale (e.g., not use the lowest or highest part of the scale), the data was normalized so that data could be pooled across the subjects. This was accomplished by converting the values to z-scores, for each subject separately. (Note that lower z-scores of the perceived boundary strength mean that the boundary was perceived as stronger, and higher z-scores mean that the boundary was perceived as weaker.)

2.3.4. Statistical analysis

A three-factor repeated measures ANOVA (within-subject factors: speaker, consonant and boundary, with each perception subject providing one averaged score per condition) was conducted on the z-scores using the statistical software package Statview (produced by the SAS Institute). Criterial significance was set at $p < .05$. All and only statistically significant results are reported.

2.4. Results: perception

The results show that listeners perceive the two IP boundaries produced by the speakers as boundaries of different strengths. For the hypothesis examined, the critical result was that there was an *interaction* of speaker and boundary ($F(3,24) = 21.384$, $p < .0001$), showing that boundary A was perceived as stronger for sentences produced by speakers K and M and that boundary B was perceived as stronger for sentences produced by speakers D and J, as shown in Fig. 2. This interaction crucially shows that the individual speaker differences in prosodic phrasing that we observed in the production part of the study are reflected in the listeners' perception of boundary strength.

In addition to this result, there were effects of boundary and an interaction effect between speaker and consonant. The effect of boundary ($F(1,24) = 4.379$, $p = .0471$) showed that overall boundary A was perceived as stronger than boundary B (in z-scores and standard deviations, boundary A = −.113 (.99), boundary B = .112 (.992)). The interaction effect of speaker and consonant ($F(3,24) = 6.856$, $p = .0004$) showed that for speaker J the boundaries where
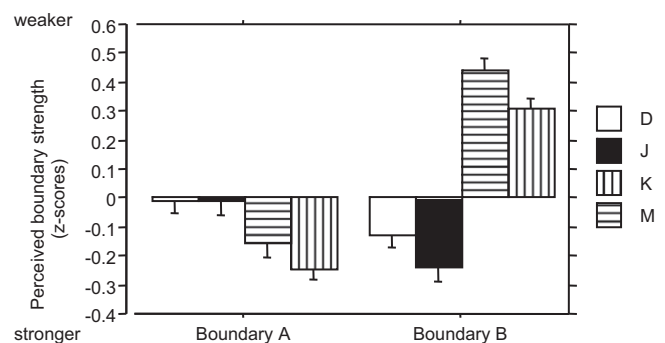

Fig. 2. Interaction of boundary and speaker effects on boundary strength perception, means (z-scores) and standard errors.

C1 was [s] were perceived as stronger, while for the other speakers the boundaries where C1 was [t] or [d] were perceived as stronger. This again corresponds to the results in production where the significant effects of consonant were such that the boundary was stronger for [t/d] than for [s] for subjects K, D, and M (there were no articulatory data for subject J).

While there were no articulatory data for subject J, the acoustic duration of the boundary for this subject could be examined (in addition to conducting the prosodic labeling, see Section 2.1.4) in order to evaluate whether the listeners' perceptual judgments reflect the speaker's prosodic phrasing. The acoustic boundary duration for sentences 1 and 2 was the duration from the burst of the preboundary stop to the burst of the postboundary stop. For sentences 3 and 4 the boundary duration was labeled from the preboundary vowel onset (as indicated by the onset of the formant structure) to the burst of the postboundary consonant. These measurements corresponded closely to the articulatory measurements. A two-factor ANOVA tested the effect of boundary (with the levels 'boundary A' and 'boundary B') and consonant (with two levels: 'alveolar stop [t]/[d]' and '[s]') on the acoustic boundary duration. The results show an effect of consonant ($F(1, 36) = 11.553$, $p = .0017$), such that [t]/[d] boundaries are longer than [s] boundaries, and an effect of boundary ($F(1, 36) = 12.261$, $p = .0013$), such that boundary B is stronger than boundary A. Thus the listeners' perception tracks the production (as seen in the acoustic measure) for this subject as well.

Summarizing the results of the perception study, listeners' judgments indicate that the two IP boundaries are perceived as being of different strength. Importantly, the results show that the perception of prosodic boundaries reflects the particular production of the speakers, in that the asymmetries observed in speakers' individual boundary strength productions (i.e., whether boundary A or boundary B was stronger) are reflected in the perception of the listeners.

3. Discussion

We investigate whether speakers produce prosodic boundaries of the same general type with different strengths, and whether listeners are perceptually sensitive to these differences in juncture strength in a way that mirrors articulation. In the production study all three speakers produce IP boundaries of different strength. In the perception study, speakers' production is reflected in the evaluation of boundary strength by listeners. Listeners perceive boundary A to be stronger than boundary B for exactly the two subjects producing boundary A as stronger than boundary B; while for the subject who produces boundary B as stronger, listeners perceive boundary B as stronger. The results from the perception of the fourth speaker also show that listeners perceive boundaries of the same prosodic type to be of different strength.

Summarizing, the results of both articulatory and perceptual analyses confirm that speakers do grade major/large/IP junctures in strength, and that listeners' assessment of disjuncture track these differences. The close connection between the production and perception of prosodic boundary strength also shows that listeners are able to perceive subtle differences choreographed in the articulatory kinematics.

To examine this connection between the production and perception of prosodic boundary strength further, we evaluate the articulatory cues that listeners might have used in the perception task.

To date, only a few studies have examined the relationship between the phonetic properties of production and the perception of prosodic boundaries, and they have focused on acoustic data. From these studies, it has emerged that the salient cues to boundary perception are pause duration, pitch reset, and final lengthening (Gussenhoven & Rietveld, 1992; Hanson, 2003; Sanderman & Collier, 1995; Swerts, 1997; Wightman et al., 1992). However, apart from

Table 4
Results of linear regression fitting *z*-scored production variables to *z*-scored perceived boundary strength values.

C1 opening movement duration	$R^2 = .035$, standardized coefficient $= -.187$, $F(1,3024) = 109.435$, $p < .0001$
C1-to-C2 peakvel	$R^2 = .061$ standardized coefficient $= -.246$ $F(1,3024) = 195.441$, $p < .0001$
V1-to-V2 peakvel (for sentences 3 and 4)	$R^2 = .063$, standardized coefficient $= -.251$, $F(1,1774) = 118.978$, $p < .0001$

two studies (Cole, Goldstein, Katsika, Mo, Nava, & Tiede, 2008; Krivokapić, 2007b), to the best of our knowledge no research has been conducted examining how perception is related to articulatory properties of prosodic boundaries. Both of these studies examined a much broader range of boundary strengths. Our study allows us to isolate articulatory correlates of small, yet reliably perceptible, boundary strength differences.

A linear regression analysis fitted the following production variables to *z*-scored perceived boundary strength values (the values of the production variables were also converted to *z*-scores, for each subject separately): C1 opening movement duration, C1 opening movement acceleration duration, C2 closing movement deceleration duration, C2 closing movement duration, C1-to-C2 peakvel, and V1-to-V2 peakvel. The significant results are shown in Table 4. Note that subjects were asked to judge how strongly connected two words are. Words with a very weak or no boundary are strongly connected and thus receive a high score in the perception experiment, and weakly connected words—with strong boundaries—received a low score. Therefore a negative correlation means that boundaries perceived as strong (low *z*-scores) correspond to longer duration of the production variable. A positive correlation means that boundaries perceived as strong (low *z*-scores) correspond to shorter duration of the production variable.

The first point to note is that the examined temporal variables are not strongly correlated to the perceived boundary strength values. This indicates that in addition to the temporal variables, tonal properties play a significant role. The role of syntactic structure on the other hand is most likely not particularly relevant in this case since different speakers produced identical sequences of words with different prosodic structures, indicating that the listeners were sensitive to the specific productions, rather than to the syntactic structure.

The results, based on R^2, show that the boundary measures (V1-to-V2 peakvel, C1-to-C2 peakvel) are the best predictors of boundary strength perception, followed by the pre-boundary opening movement duration. The responsiveness of listeners to the pre-boundary opening movement corroborates previous findings (Cole et al., 2008; Krivokapić, 2007b). It also relates to what is known about boundaries from acoustic and articulatory studies, namely that the most reliable temporal indicator of prosodic boundaries is found in pre-boundary lengthening, typically examined in the rhyme of the final syllable, and in the articulatory movement closest to the boundary. However, the implication of the results presented here is that the dominant indicators of boundaries, as perceived by listeners, are those parameters that *span* the boundary and *extend* over a period of time, rather than those parameters that correspond to one constriction/release movement or one part of one movement.

It is interesting to observe that listeners do not seem to be responsive to variables that presented inconsistent results in the production study: In evaluating the production variables it was noted that, while overall subjects produced either boundary A or boundary B as stronger, there were some results pointing in opposing directions (e.g., C1 acceleration duration for subjects K and A, and C2 deceleration duration for D had an opposite trend

than other measures of boundary strength for these subjects). Generally, these variables didn't show significant results in the regression analysis. On the assumption that listeners are responsive to the most relevant and stable production variables, the results of the linear regression analysis lend further support for the findings that the observed boundary strength differences are an objective description of the data, despite some production variables behaving differently.

In recent years, the role of prosodic phrasing in speech processing has become increasingly apparent. On the production side, it has become clear that prosodic boundaries modulate the articulation of gestures in systematic ways. It terms of function, it has been argued that speakers use prosodic phrasing as a way to make their processing load easier in production (Frazier, Carlson, & Clifton, 2006; Krivokapić, 2007a,b). It has also been noted that prosodic structure is relevant in listener comprehension (see e.g., Cutler, Dahan, & Donselaar, 1997 for an overview); prosodic structure has been argued to be guiding listeners in comprehension in cases of syntactic ambiguity (e.g., Carlson, Clifton, & Frazier, 2001; Kjelgaard & Speer, 1999; Jun, 2003; Schafer, 1997) and lexical ambiguity (Cho, McQueen, & Cox, 2007; Salverda, Dahan, & McQueen, 2003). Studies examining phrase initial segments have found effects of prosodic boundaries in word recognition (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; McQueen & Cho, 2003). These studies show that prosodic structure is an essential part of the grammatical structure of language that is used by both listeners and speakers in language processing. The present study is consistent with these findings—speakers exhibit a fine-grained distinction between prosodic boundaries that listeners in turn are sensitive to.

This experiment adds to the relatively small but consistent body of evidence showing that gross distinctions among prosodic phrase types must flexibly allow for more fine-grained distinctions that speakers deploy and that listeners are sensitive to. As articulatory studies become more available to examine prosodically rich natural discourse as well as read speech in laboratory studies, prosodic variation beyond pitch tracking and acoustic rhyme duration will increasingly provide direct evidence as to communicative prosodic structuring. Such articulatory variation is likely to be exploited in a variety of situations by talkers. There is every reason to think that boundaries are created and instantiated over an interval in which spatiotemporal deformations occur in speech as informational groups are constructed and implemented by a speaker (Byrd & Saltzman, 2003; Ferreira, 1991,1993; Keating & Shattuck-Hufnagel, 2002; Krivokapić, 2007a,b). Furthermore, the alignment of speakers and listeners in the joint creation of prosody in dialog exists at the phonetic level (e.g., Kim and Nam, 2009; Krivokapić, 2010; Smith, 2007) and is a rich area for further exploration. The fine differences in boundary strength that speakers produce and listeners perceive might be critical to how alignment is established, possibly as part of a multimodal approach (e.g., articulatory, acoustic, manual, gaze, head-movement) to the mutual generation of prosody.

3.1. Theoretical implications of treatment of prosodic structure

Within an approach utilizing a limited number of prosodic categories (word, ip, and IP), the most likely structural explanation of the result above—that both of the boundaries are of the IP category type but still differ in strength—is that the difference in strength reflects two levels of embedding of the same category (shown in the left and right part of Fig. 3). In other words, the explanation could be that these are *recursive* prosodic structures. Assuming a structure nesting like categories can account for the results from both the perception and production study while avoiding assuming a new category. For arguments for recursion, see for example Dresher (1994), Ladd (1986), Itô and Mester (1992, 2010),
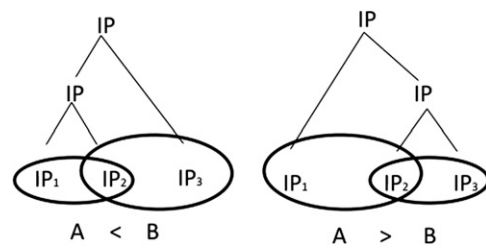

Fig. 3. Prosodic structures for sentences 1, 2, 3, and 4.

Kubozono (1992), Selkirk (1995), McCarthy and Prince (1993), Wagner (2010), and Schreuder, Gilbers and Quené (2009).

The results of the experiment could also be accounted for if an *additional* prosodic level above the IP level is assumed (given the specific phonetic correlates of the boundaries in this study, a level below the IP for these data is unlikely). Different theories have postulated such additional levels (e.g., Hayes, 1989; Nespor & Vogel, 1986; Selkirk, 1986; for an overview see Shattuck-Hufnagel & Turk, 1996). Below we examine the type of evidence usually presented for motivating prosodic levels, in order to try to establish whether the results presented here are better accounted for by assuming an additional prosodic category or by assuming recursion. The problematic nature of the question of how many prosodic categories should be assumed, and arguments against a proliferation of categories, are discussed in detail in Ladd (1996), Itô and Mester (2010:147f), Selkirk (2009:37f), Wagner (2010), and Tokizaki (2002).

Two kinds of evidence are generally brought forward to argue for the existence of a particular level: the existence of a phonological rule that applies only at a specific level of the prosodic hierarchy or the existence of phonetic correlates of a prosodic category. The first type of evidence requires that there are rules that apply only at a specific level of the prosodic hierarchy, but not at the prosodic levels above or below that. No such rules are known to exist however (Byrd, 2006; Wagner, 2005). Claims about such rules have been made based on impressionistic evidence but have not been empirically validated.[4]

The second type of evidence consists of specific phonetic properties (including temporal and tonal properties) signaling a specific prosodic boundary. Crucially, these properties would need to differ in a categorical manner in order to signal different prosodic category types (e.g., Frota, 2000; Wagner, 2005). Thus for example, as outlined above, ip and IP differ not just in degree of final lengthening but also in the existence of a boundary tone at the IP but not at the ip level.

Categories above the IP level that, based on their phonetic properties, could potentially account for the data in this study have been suggested in the literature. Price et al. (1991) suggest two more levels above the level of IP (category 5 and category 6; category

---

[4] One of the rules that has been examined is *raddoppiamento*, a lengthening rule generally thought to apply at the phonological phrase and thus thought to motivate it. Experimental evidence however showed that *raddoppiamento* applies at the IP level as well, even when a pause exists at the boundary (Campos-Astorkiza, 2004). Similarly, *s*-voicing in Greek has been argued to apply across prosodic domains up to the IP level, but not across IP boundaries, and this has been taken to be an argument for assuming the IP phrase (Nespor & Vogel, 1986). However, *s*-voicing has been found to apply in a gradient manner at word and ip boundaries, with some consonants not being voiced at all, others being partially voiced, and others still becoming fully voiced (Pelekanou & Arvaniti, 2002). The *s*-voicing phenomenon is thus more likely the result of gestural overlap, rather than a phonological rule applying obligatorily and defining a prosodic category (Pelekanou & Arvaniti, 2002). In Jun's (2003) study of the prosodic structure in Korean, one observes a number of phonological phenomena that apply up to a prosodic category, but none that apply only at a certain type of boundary, and not above or below it as well. Thus to date there does not seem to be clear evidence of rules that would support a specific number of prosodic categories and that would either justify or argue against additional levels of the prosodic hierarchy.

4 being the IP in the ToBI model). Wightman et al. (1992), in their study finding a larger number of distinctions in category strength than other studies, show that one of their measures, namely the duration of the coda consonant closest to the boundary, seems to distinguish between the IP and a category above it, category 5. However, note that this higher-than-IP category would be defined by a stronger percept of boundary strength and by more acoustic lengthening, thus relying on continuous characteristics and without a distinctive property marking it. This lack of categorical differentiation between an IP and potential higher categories is one of the reasons such higher levels have been rejected in ToBI as part of the prosodic hierarchy (Beckman, Hirschberg, Shattuck-Hufnagel, 2005), even though the number of categories in ToBI has its origins in the findings reported in Price et al. (1991) and Wightman et al. (1992). This "super" category then does not seem to provide a good explanation of the data presented here.

Another type of category has been suggested by Beckman et al. (2005). They propose that more categories can be distinguished above the IP level, but that these are of a qualitatively different nature from the traditional phonological categories such as IP and ip since they are distinguished by gradient properties and are potentially recursive. The dominating IP in our data then could in principle be argued to fall under this characterization. Note though that these higher categories in Price et al. (1991), Wightman et al. (1992) and Beckman et al. (2005) are argued specifically to occur in long sentences, or marking the sentence end, or a discourse 'paragraph.' Price et al. (1991:2962) note that break index 5 is "a boundary marking a grouping of intonational phrases" and "[a] break index of 5 is typically found in long sentences and frequently coincides with a breath intake or long pause," and the same definition is used in Wightman et al. (1992). Break index 5 would correspond to the first category above the IP (which is marked with break index 4). Break index 6 is used to mark sentence boundaries. Crucially, both of these are unlikely to correspond to the boundaries examined in this study, where the target junctures are not utterance final boundaries and the sentences used in the experiment are very short. Thus IP recursion might be a better way to structurally account for the findings in this study.

To summarize: The results of our study can lend support to two structural possibilities: (a) the IP category is recursive, or (b) the prosodic hierarchy has more categories above the prosodic word level than ip and IP. For theories that assume only two prosodic levels above the word, a recursive structure is a viable explanation of the data as it accounts for the results without assuming a new category. For theories that assume more than two prosodic categories above the word, the discussion above indicates that there does not seem to be independent evidence for additional categories above the IP that would correspond to the data in this specific study. Nevertheless, the possibility of an additional category above the IP cannot be excluded.

4. Conclusion

We have presented here two complementary studies of prosodic boundary strength. The experimental work investigates (1) whether articulatory gestures at large/IP boundaries can exhibit temporal distinctions that indicate differences in boundary strength and (2) to what extent listeners' judgments reflect specific and detailed differences among boundaries. We combine perceptual and articulatory experimentation to evaluate these questions. Kinematic data show that speakers produce IP boundaries of differing strengths. Listeners perceive the differences in prosodic boundary strength in a manner that aligns with speakers' specific production. The experiments provide support for the existence of prosodic strength differences among IP boundaries and demonstrate a close link between the articulation and perception of prosodic boundaries.

Acknowledgments

Appendix I

Instruction sheet for sentence 1:
During the experiment, please read the question silently, and the answers aloud. **<u>Please read carefully, paying attention to punctuation.</u>**
Below are the sentences you will be reading. Please familiarize yourself with the question and answer. You can do so by reading them aloud or to yourself.
What is the workers' shift?

THEY USUALLY DO: ᴛᴏ 2:00, from 8:00.

Instruction sheet for sentence 4:

During the experiment, please read the question silently, and the answers aloud. **<u>Please read carefully, paying attention to punctuation</u>**.
Below are the sentences you will be reading. Please familiarize yourself with the question and answer. You can do so by reading them aloud or to yourself.
What range on the map grid do they have?
THEY USUALLY SEE: ᴛᴏ 2, from C.

Appendix II

The experiment consists of several parts. In each part of the experiment, you will see sentences in front of you. You will be asked to read them. Each part will be preceded by instructions.

PRACTICE:
The following sentences are similar to the ones that you will be reading during the experiment. For each part: please read the instructions, and then read the answers aloud to the experimenter.

- -

A
Instructions:
1. Please read the question silently, and the answers aloud. <u>Please read carefully, paying attention to punctuation.</u>
SENTENCES:
What is their parents' shift?
   1. THEY GENERALLY DO: TO 5:00, from 11:00.
   2. THEY GENERALLY DO: TO 7:00, from 3:00.
   3. THEY GENERALLY DO: TO 6:00, from 4:00.

- -

B
INSTRUCTIONS:
2. Please read the question silently, and the answers aloud. <u>Please read carefully, paying attention to punctuation.</u>
SENTENCES:
What is their parents' shift?
   4. THEY GENERALLY DO: FROM 11:00, to 5:00.
   5. THEY GENERALLY DO: FROM 3:00, to 7:00.
   6. THEY GENERALLY DO: FROM 4:00, to 6:00.

- -

C
INSTRUCTIONS:
Please read the question silently, and the answers aloud. <u>Please read carefully, and do *not* pause within the sentence.</u>
SENTENCES:
What is their parents' shift?
   7. 11:00 to 5:00.
   8. 3:00 to 7:00.
   9. 4:00 to 6:00.

- -

D
INSTRUCTIONS:
Please read the question silently, and the answers aloud. <u>Please read carefully, paying attention to punctuation.</u>
SENTENCES:
What range on the map grid do they have?
   10. THEY GENERALLY HAVE: TO 5, from A.
   11. THEY GENERALLY HAVE: TO 1, from H.
   12. THEY GENERALLY HAVE: TO 4, from B.

- -

E
INSTRUCTIONS:
Please read the question silently, and the answers aloud. <u>Please read carefully, paying attention to punctuation.</u>
SENTENCES:
What range on the map grid do they have?
   13. THEY GENERALLY HAVE: FROM A, to 5.
   14. THEY GENERALLY HAVE: FROM H, to 1.
   15. THEY GENERALLY HAVE: FROM B, to 4.

- -

F

INSTRUCTIONS:

Please read the question silently, and the answers aloud. Please pay attention to punctuation, and do *not* pause within the sentence.

SENTENCES:

What range on the map grid do they have?

16. A to 5.
17. H to 1.
18. B to 4.

The experiment consists of several parts, and is very similar to the practice you have just had. In each part of the experiment, you will see sentences in front of you. You will be asked to read them. Each part will be preceded by instructions. During the experiment, start reading each sentence when you hear a beep sound. You will need to wait a moment between each sentence because when you finish reading there will be one beep to signal the end of the sentence, and then the start beep to signal you to read the next sentence. Please also *sit still* and try not to move your head while you are speaking. Do of course let us know if you become uncomfortable or if you feel that any of the sensors might be becoming loose.

References

Beckman, M. E., & Ayers Elam, G. A. (1997). Guidelines for ToBI labelling. Version 3.0, unpublished ms. Available online at: ⟨http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf⟩.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In: S.-A. Jun (Ed.), *Prosodic typology. The phonology of intonation and phrasing* (pp. 9–54). Oxford University Press.

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

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

Byrd, D. (2006). Relating prosody and dynamic events: Commentary on the papers by Cho, Navas, and Smiljanić. In: L. Goldstein, D. H. Whalen, & C. Best. (Eds.), *Laboratory phonology 8: Varieties of phonological competence* (pp. 549–561). Mouton de Gruyter.

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In: M. B. Broe, & J. B. Pierrehumbert. (Eds.), *Papers in laboratory phonology 5. Acquisition and the lexicon* (pp. 70–87). Cambridge University Press.

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., & 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, 149–180.

Campione, E., & Veronis, J. (2002). A large-scale multilingual study of silent pause duration. In: Bernard Bel, & Isabelle Marlien (Eds.), *Proceedings of speech prosody 2002, 11-13 April 2002* (pp. 199–202). Aix-en-Provence: Laboratoire Parole et Langage.

Campos-Astorkiza, R. (2004). Italian Raddoppiamento: Prosodic effects on length. 148th Meeting of the Acoustical Society of America, San Diego, CA, November 2004. *Journal of the Acoustical Society of America*, 116(4), 2645.

Carlson, K., Clifton, C., Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45, 58–81.

Cho, T. (2006). Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. In: L. Goldstein, D. H. Whalen, & C. Best. (Eds.), *Laboratory phonology 8: Varieties of phonological competence* (pp. 519–548). Mouton de Gruyter.

Cho, T., McQueen, J., & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35, 210–243.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler., J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language*, 51, 523–547.

Cole, J., Goldstein, L., Katsika, A., Mo, Y., Nava, E., & Tiede, M. (2008). Perceived prosody: Phonetic bases of prominence and boundaries. Acoustical Society of America, Miami, FL, November 2008. *Journal of the Acoustical Society of America*, 124(44), 2496.

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

Cutler, A., Dahan, D., & Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.

Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619.

Dresher, E. B. (1994). The prosodic basis of the Tiberian Hebrew system of accents. *Language*, 70, 1–52.

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

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30, 210–233.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 233–253.

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

Frazier, L., Carlson, K., & Clifton., C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10, 244–249.

Frota, S. (2000). *Prosody and focus in European Portuguese. Phonological Phrasing and Intonation*. Garland Publishing.

Fuchs, S., Perrier, P., Geng, C., & Mooshammer, C. (2006). What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents. In: J. Harrington, & M. Tabain. (Eds.), *Speech production: Models, phonetic processes and techniques* (pp. 149–164). New York: Psychology Press.

Gaitenby, J. H. (1965). The elastic word. *Haskins report* SR-2, 3.1-3.12.

Granqvist, S. (1996). Enhancements to the Visual Analogue Scale, VAS, for listening tests. *TMH-QPSR, 4/1996, 61-62*. Department of Speech, Music and Hearing, Royal Institute of Technology.

Gussenhoven, C., & Rietveld., A. C. M. (1992). Intonation contours, prosodic structure, and preboundary lengthening. *Journal of Phonetics*, 20, 283–303.

Hansson, P. (2003). *Prosodic phrasing in spontaneous Swedish*. Doctoral Dissertation. Travaux de l'institut de linguistique de Lund 43. Lund: Department of Linguistics and Phonetics, Lund University.

Hayes, B. (1989). The prosodic hierarchy in meter. In: P. Kiparsky, & G. Youmans (Eds.), *Rhythm and meter* (pp. 201–260). Academic Press.

Itô, J., & Mester, A. (1992). *Weak layering and word binarity. Linguistics Research Center-92-09*. Santa Cruz: University of California.

Itô, J., & Mester, A. (2010). The extended prosodic word. In: B. Kabak, & J. Grijzenhout (Eds.), *Phonological domains: Universals and derivations* (pp. 135–194). Mouton de Gruyter.

Jun, S.-A. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research*, 32, 219–249.

Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA working papers in phonetics*, 101, 112–156.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and inference in the resolution of temporal syntactic closure ambiguity. *Journal of Memory and Language*, 40, 153–194.

Kim, M., & Nam, H. (2009). Pitch accommodation in synchronous speech. *Journal of the Acoustical Society of America*, 125(4), 2575.

Klatt, D. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics*, 3, 129–140.

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

Krivokapić, J. (2007b). *The planning, production, and perception of prosodic structure*. Doctoral Dissertation, University of Southern California, Los Angeles, CA.

Krivokapić., J., & Ananthakrishnan, S. (2007). Gradiency and categoricity in prosodic boundary production and perception. Acoustical Society of America, New Orleans, LA, November 2007. *Journal of the Acoustical Society of America*, *122*(5), 3020.

Krivokapić, J. (2010). Prosodic interaction between speakers of American and British English. Acoustical Society of America, Baltimore, MD, April 2010. *Journal of the Acoustical Society of America*, *127*(3), 1851.

Kubozono, H. (1992). Modeling syntactic effects on downstep in Japanese. In: G. Docherty, & R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 287–368). Cambridge University Press.

Ladd, D. R. (1986). Intonational phrasing: The case of recursive prosodic structure. *Phonology Yearbook*, *3*, 311–340.

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. (1996). *Intonational phonology. Cambridge Studies in Linguistics*. Cambridge University Press.

McCarthy, J., & Prince, A. (1993). Generalized alignment. *Yearbook of Morphology*, *1993*, 79–153.

McQueen, J. M., & Cho., T. (2003). The use of domain-initial strengthening in segmentation of continuous English speech. In: M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 2993–2996). Adelaide: Casual Productions.

Mooshammer, C., Hoole, P., & Geumann, A. (2006). Jaw and order. *Language and Speech*, *50*, 145–176.

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

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.

Pelekanou, T., & Arvaniti, A. (2002). *Postlexical rules and gestural overlap in a Greek spoken corpus. Recherches en linguistique grecque*, Vol. I. Paris: L' Harmattan pp. 71–74.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Doctoral Dissertation, MIT Press. Distributed 1988, Indiana University Linguistics Club.

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.

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.

Rietveld, T., & Chen., A. (2006). How to obtain and process perceptual judgements of intonational meaning. In: S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schliesser (Eds.), *Methods in empirical prosody research* (pp. 283–319). Berlin/New York: Mouton De Gruyter.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.

Sanderman, A. A., & Collier, R. (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: American Institute of Physics.

Schafer, A. J. (1997). *Prosodic parsing: the role of prosody in sentence comprehension*. Doctoral Dissertation, University of Massachusetts, Amherst, Massachusetts.

Schreuder, M., Gilbers, D., & Quené, H. (2009). Recursion in phonology. *Lingua*, *119*, 1243–1252.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–405.

Selkirk, E. (1995). The prosodic structure of function words. In: J. Beckman, L. Walsh Dickey, & S. Urbanczyk. (Eds.), *Papers in optimality theory* (pp. 439–470). GLSA Publications.

Selkirk, E. (2009). On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Gengo Kenkyu*, *136*, 35–73.

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

Smith, C. L. (2007). Prosodic accommodation by French speakers to a non-native interlocutor. In: *Proceedings of the 16th international congress of phonetic sciences*, 1081–1084.

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

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experiment Psychology: General*, 64–81.

Tabain, M. (2003). Effects of prosodic boundary on/aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America*, *113*, 2834–2849.

Tokizaki, H. (2002). Prosodic hierarchy and prosodic boundary. *Bulletin of Universities and Institutes*, *56*, 81–99.

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

Wagner, M. (2005). *Prosody and recursion*. Doctoral Dissertation, Massachusetts Institute of Technology.

Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language and Linguistic Theory*, *28*, 183–237.

Wagner, M., & Crivellaro., S. (2010). Relative prosodic boundary strength and prior bias in disambiguation. *Speech Prosody*, *2010* 100238:1–4.

Wewers, M. E., & Lowe, N. K. (1990). A critical review of Visual Analogue Scales in the measurement of clinical phenomena. *Research in Nursing & Health*, *13*, 227–236.

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.

Yoon, T. -J. (2007). *A predictive model of prosody through grammatical interface: A computational approach*. Doctoral Dissertation, University of Illinois at Urbana-Champaign.