Effects of dynamic pitch and relative scaling on the perception of duration and prosodic grouping in American English

Alejna Brugos and Jonathan Barnes

Boston University, Boston, Massachusetts, USA

abrugos@bu.edu

Abstract

Results of two perception experiments suggest that using timing measures alone to compute prosodic structure misses valuable information from pitch. Previous research showed that pitch can distort perceived duration: tokens with dynamic or higher f0 are perceived as longer than comparable level-f0 or lower-f0 tokens, and silent intervals bounded by tokens of widely differing pitch are heard as longer than those bounded by tokens closer in pitch (the kappa effect). Phrase edges (signalled by increased duration, pause, phrase tones, and f0 reset) set the scene for pitch to modulate perceived duration. Two new experiments used the same duration and f0 manipulations (level vs. varying-slope rises, at varying pitch ranges) of segmentally-identical base files, in two separate tasks: 1) a linguistic grouping task using an ambiguously-structured phrase and 2) a psychoacoustic study on perceived duration. Results show that effects on perceived duration due to dynamic pitch can be either strengthened or nullified depending on relative scaling of compared tokens. These same manipulations push grouping judgments beyond what would be expected from distortions of perceived duration. This suggests that listeners integrate pitch and timing cues when judging linguistic structure, supporting measures of relative boundary size that combine duration and pitch measures.

Index Terms: duration perception, auditory illusions, dynamic pitch, timing, prosodic grouping, boundary tones

1. Introduction

Edges of prosodic groups are known to be marked (at least in many languages, Fon (2012) [1]) by pre-boundary lengthening, silent pauses, phrase tones and reset. Phonetic measures of these features are typically taken independently, without consideration of how they may interact in perception. However, perception of time can be systematically affected by a range of contextual factors (Brown, 2008) [2], including pitch (Hoopen, 2008) [3]. A growing body of work from a diverse range of fields shows that pitch and timing are not entirely perceptually independent.

1.1. Pitch-time interaction

Since Lehiste (1976) [4] showed that subjects perceived vowels with dynamic f0 as longer in duration than static-f0 vowels of the same objective duration, many studies have tried to replicate this finding, with varying results (see Cumming, 2011 [5] for an overview). Cumming (2011) [5] and Yu (2010) [6] both reproduced this effect, with differing methodologies and languages; Yu also found that vowels with higher f0 were perceived as longer than lower-f0 vowels.

Outside of speech research, there is a substantial body of experimental work on pitch-timing interaction in perception. Henry (2011) [7] showed that perception of duration of non-speech tone glides can be modulated by the pitch change

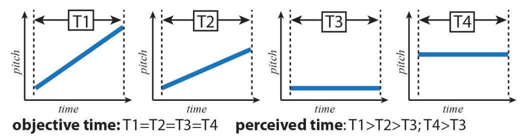

Figure 1: *Schematic of dynamic pitch & scaling effects on perceived duration of filled intervals. The intervals (T1, T2, T3 & T4) are objectively equal in duration; those with dynamic pitch sound longer than level, and more so with steeper pitch. Higher pitch intervals are also perceived as longer than those with lower pitch.*

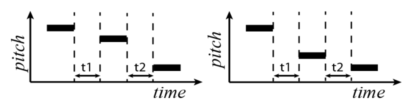

Figure 2: *Schematic example of the auditory kappa effect, whereby relative tone height affects timing perception. Silent intervals (t1, t2) are equal, but t1 sounds shorter at left, longer at right.*

velocity of the target glide and of the standards: glides with greater pitch change velocity are perceived as longer than those with lesser pitch change velocity, or level-pitch tones.

Other work has focused on the effects on perceived duration of silent intervals bounded by filled intervals of varying pitch distance, a phenomenon known as the auditory kappa effect: Silent intervals bounded by tones of closer pitch proximity are perceived as shorter in time than those of equal objective duration bounded by tones of a greater pitch distance (Cohen et al., 1953[8], 1954[9]). While typically demonstrated using non-speech tones (Shigeno, 1993 [10], Crowder & Neath, 1995 [11]; MacKenzie, 2007 [12]; *inter alia*), Brugos & Barnes (2012b) [13] showed that the auditory kappa effect also obtains for spoken language, such that the perceived duration of silent pauses in speech was modulated by the pitch distance across those pauses. In a second study using identical materials, Brugos & Barnes (2012a) [14] found that the effect of these pitch manipulations was even greater on perceived prosodic grouping of these phrases: even effects of objective duration differences on grouping perception were in some cases overridden. These results suggest that relative pitch proximity of neighboring prosodic phrases, described in the literature as phrase-initial reset (Jun, 2003) [15], should be taken into account for estimating boundary size, and support a trading relationship between pitch and timing cues in prosodic grouping (Beach, 1991 [16]; Cumming, 2011b [17]; Jeon & Nolan, 2013 [18]).

Of course, since phrase boundaries in natural speech commonly play host to dynamic pitch (e.g., boundary tones), pitch jumps (e.g., reset), and durational variation (phrase-final lengthening/pauses), we might expect all these pitch and timing cues to enter into cue-trading relationships for the signalling of prosodic grouping. In fact, something of this sort has been shown in a variety of studies. When f0 cues are neutral, grouping can be cued by duration cues alone (Scott, 1982 [19]; Wagner & Crivellaro, 2010 [20]); likewise f0 cues alone can cue grouping when pitch cues are held constant (House, 1990 [21]). When pitch and timing are manipulated together, the picture of cue interactions for prosodic grouping becomes more complex (Beach, 1991 [16]; Jeon & Nolan, 2013 [18], 2010 [22]; Cumming, 2011b [17]).

2. Two experiments

To investigate how dynamic pitch and duration might interact in both duration and grouping perception in American English, a pair of experiments was designed using the same stimuli in two different tasks: 1) a linguistic judgment of grouping in an ambiguous phrase based on cue interpretation and 2) a psychoacoustic judgment of perceived duration.

2.1. Methods

The ambiguous phrase chosen as context for the linguistic grouping task was a string of color terms *blue and green and purple* (following methodology of Beach et al., 1996 [23]). This phrase can be parsed variously: 1) ungrouped (a simple list of 3 colors) or 2) two groups, one pair of colors and a third color on its own, i.e.: *blue and (green and purple)* (B-GP) or *(blue and green) and purple* (BG-P).[1]

The f0 pattern of base recordings of *blue* was manipulated to include both dynamic f0 and plateau contours, and crossed with a continuum of duration manipulations leading to changes in relative duration of the words *blue* and *green*. Assuming that *blue* being longer than *green* cues more B-GP responses, and that dynamic f0 cues longer perceived duration, then we might predict dynamic F0 in *blue* likewise to cue more B-GP responses: results of duration and grouping perception tasks, in other words, should be largely overlapping. However, if, like in Brugos & Barnes 2012a [14], the effects of pitch go beyond their modulation of perceived duration, results from the two tasks are expected to diverge.

2.1.1. Stimuli

Manipulations of pitch and duration were performed to an identical base recording of *blue*, and these same resultant resyntheses were used in both the grouping perception and duration perception tasks as described below. A single version of the complete phrase was used as a base file for additional manipulations of f0 and duration to produce stimuli in the experiments. In order to see whether pitch manipulations modulate grouping perception by way of timing perception, it was necessary first to create a neutral condition in which timing manipulations alone might shift perceived grouping. A female native speaker of American English (the first author) produced multiple versions of the phrase, and the eventual base file token was selected through an extensive process of evaluation, resynthesis, and concatenation of naturally spoken words. Durations and intensity of each of the words were adjusted to produce a natural-sounding concatenation that did not strongly cue either B-GP or BG-P grouping. Base durations and duration continuum points were chosen based on a pre-experiment screening with 12 subjects via web form.

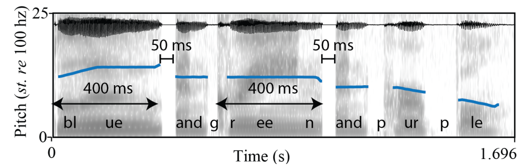

Figure 3: *The base/neutral file. The words* blue *and* green *are ~ 400* ms *long. The f0 contour of* blue *is a plateau with a 2* st *rise, and then level f0,* and green *is level f0, 2* st *below the max f0 of blue,* and purple *starts another 2* st *lower, and ends in a 4* st *fall.*

Pitch and timing manipulations: The base recording of *blue* was resynthesized to create 3 f0 contour shapes at 5 durations, and at 3 steps affecting f0 range. All tokens of blue began with a rise similar to what was seen in natural productions: the 2 *st* rise began at the onset of voicing, through the [l] and into the beginning of the vowel [u]. In order to reduce segmental variation in the onset that might cue differences in perceived prominence and grouping, all manipulations for duration and contour were done only to the /u/ portion of the word following this pivot point (at 158 milliseconds into the word). From this point, the f0 did one of 3 things: 1) stayed level to the end of the word ("plateau"), 2) rose an additional 2 *st* to the end of the word ("2-st-rise") or 3) rose 4 *st* from the pivot ("4-st-rise"). 5 duration manipulations were performed on this same post-pivot interval such that the total duration of the word equalled 300 *ms*, 350 *ms*, 400 *ms*, 450 *ms*, and 500 *ms*, creating 15 time-by-contour manipulations (Figure 4, left).

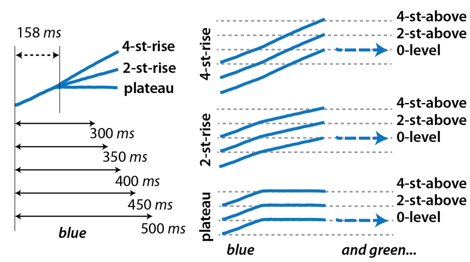

Figure 4*: The base file* blue *was resynthesized to create 3 contour shapes at 5 durations, and duration of the initial 2* st *rise was held constant (left). F0 contours for* blue *were each shifted to 3 f0 steps.*

Unfortunately, introducing a comparison of dynamic vs. static pitch into an experiment such as ours turns out to be far from simple. As Figure 5 shows, any attempt to alter f0 dynamicity during the word *blue* in our sequence necessarily alters the pitch gap across the following boundary also. Given the results of Brugos & Barnes (2012b)[14], this turns out to be a serious potential confound for any investigation of the effects of dynamic pitch on perceived duration.

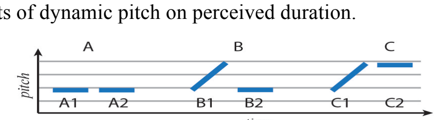

Figure 5: *Schematic showing how dynamic f0 can introduce pitch differences across phrases.*

[1] It should be noted that there is controversy in the literature as to whether it is primarily just relative boundary size, rather than categorical identity of the boundaries involved, that matters most for interpretation (Price et al. 1991 [24]; Clifton et al., 2002 [25]; Jun, 2003 [15]; inter alia). We remain agnostic here as to whether the prosodic phrases used in this study are instantiations of specific levels of the Prosodic Hierarchy (Selkirk, 1986) [26], or instead recursive or gradiently-sized groups (Ladd, 1986 [27]; Wagner, 2005 [28]; Shreuder, 2006 [29]; Kentner & Féry, 2013 [30]). The points we wish to make regarding the implementation of phrasing hold equally well under both scenarios.

In order to control for this confound, we chose to actively manipulate pitch step orthogonally to the manipulations of dynamic pitch so that we could better separate the effects. Each resulting contour/time combination was resynthesized at 3 different f0 ranges ("pitch steps") based on the pitch relationship between the end point of *blue* and the f0 of the immediately following words in the grouping experiment (*and green),* which were always level at 202 Hz. 3 pitch steps were chosen: 1) 2 *st* above *green* (the level of the neutral condition for the plateau contour shown in Figure 3, above), 2) ending 4 *st* above *green* and 3) ending level to *green*. In each case, the entire contour was shifted up or down in pitch space.

Because manipulations to *blue* alone were not sufficient to cue grouping differences for all listeners in a pre-experiment screening, and as it was obvious to some listeners when the post-*blue* phrase was unchanging, manipulations to the duration of *green* were also added. Base *green* was approximately 400 *ms* in duration, and additional manipulations yielded *green* at 350 and 450 *ms* as well.

2.1.2. Subjects, presentation and task

16 native speakers of American English (age 18 to 22 years) participated in Experiment 1 for a payment of $10. 9 of these subjects additionally participated in Experiment 2 for an additional $10. (Experiment 2, the duration task, was always presented to subjects after completion of Experiment 1.) Subjects faced a laptop and listened to stimuli over headphones. Both experiments were forced-choice tasks, with responses indicated via a button box or designated keys on the laptop. Each experiment took about 25 minutes, including breaks and training. Subjects read a brief introduction to the study, then proceeded to a training section, to ensure that they understood the task. For Experiment 1, subjects were presented with natural examples of prosodic grouping (of repeated digits) produced by a naïve speaker. Subjects proceeded to the experimental phase after answering at least 75% correct of at least 10 training trials. For Experiment 2, training consisted of presentations of plateau tokens of differing durations, using only duration differences of 100 *ms* or greater. Subjects proceeded to the experimental phase upon correctly answering 5 training trials in a row.

Experiment 1: Grouping perception. The screen presented 2 images representing two grouping choices: 1) One solid blue ball, and another purple with green spots (B-GP) and 2) one blue ball with green spots, and another solid purple (BG-P). The text "blue and green & purple" and "blue & green and purple" accompanied each image. For each trial, subjects were played a recording of the complete phrase *blue and green and purple*, and asked to indicate whether they heard the phrase as corresponding to B-GP or BG-P. They were not instructed to attend to any specific aspect of the signal. Trials included phrases with the 45 blue manipulations (3 contours x 5 durations x 3 time steps), paired with the phrase completion,

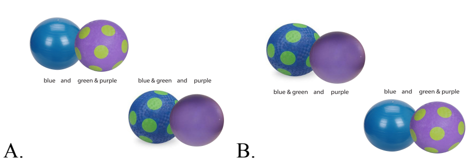

Figure 6: *Images presented to subjects for Experiment 1, with text and images suggesting the 2 groupings. Subjects either saw only A, or only B.*

including 3 durations of green (350, 400, and 450 *ms*): 4 repetitions with 400-*ms green*, and 2 repetitions each for the other 2. These 360 trials were randomized for each subject.

Experiment 2: **Duration perception**. Target versions of the word *blue* identical to those in Experiment 1 were compared to level-f0 standards of the same base file *blue*. Standards were completely level-f0 versions of *blue* at 202 Hz (the level of *and green* in the grouping experiment), and presented in the same 5 durations of the targets (300, 350, 400, 450 and 500 *ms*). Trials consisted of a target followed by a standard, with 200 *ms* of silence interceding. Each pairing was also presented in the opposite order, that is, standard followed by target (45 targets x 5 standards x 2 orders = 450 trials). An additional repetition of each of the 2 rise contours in the target-standard order was included to maximize repetitions of the comparisons of greatest interest, for a total of 600 trials, randomized for each subject. Subjects were asked to indicate which of the two repetitions of the word *blue* sounded longer.

3. Results and analysis

Experiment 1: Results presented are from 5598 experimental trials for 16 subjects. Figure 7 displays proportion of B-GP responses as a function of duration difference between the words *blue* and *green* (durations of *and* & *purple* remain constant for all conditions). Positive time values indicate that *blue* is longer than *green*, and negative that *blue* is shorter. At left, individual lines represent the 3 contours (plateau, 2-st-rise, 4-st-rise), and at right the 3 pitch steps (0-level, 2-st-above, 4-st-above). The upward diagonal trend of the lines in both graphs shows that responses are strongly correlated with duration difference between *blue* and *green*. Bigger time values show more responses that *blue* is grouped separately (B-GP), and smaller time values more responses that *blue* is grouped with *green* (BG-P). The three lines by contour (left) show virtually no separation, but the 3 lines by pitch step (right) show clear separation such that trials where the pitch of *blue* ends level with the following words show proportionately greater responses that *blue* grouped with *green* (BG-P), and higher pitch steps show increasingly more responses that *blue* is grouped separately (B-GP).

Results were analyzed using mixed-effects logistic regression, implemented through the lme4 package (Bates & Maechler, 2009 [31]) in R with response ("B-GP" or "BG-P") as dependent variable, and time step, pitch step and f0 contour as fixed factors. Subject was included as a random effect (Baayen et al., 2008 [32]). The result was a model (N = 5598, log-likelihood = -3269) showing an expected significant main effect of time step (Wald Z = 29.57, p < .001), as well as main effects for pitch step, with 2-st and 4-st differing from 0-level (2-st: Wald Z = 5.371, p < .001; 4-st, Wald Z = 7.54, p < .001). There was a weak effect of contour (between plateau and 2-st-rise (Wald Z= 1.98, p < .05), also pushing responses toward B-GP. There was also a somewhat complicated series of interactions between contour and pitch step such that dynamic pitch contours showed a slight tendency at the 2-*st* and 4-*st* steps to produce fewer B-GP responses, which will be addressed in the discussion section.

Experiment 2: Results are presented for 9 subjects for 3360 trials (only target-standard order trials are included, being closest to conditions of Experiment 1.) Figure 8 displays the proportion of "target longer" responses as a function of duration difference between target and standard.

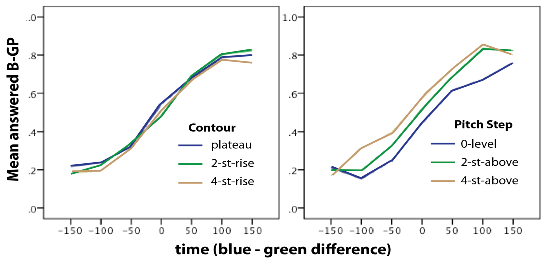

Figure 7: *Experiment 1 results: Mean responses B-GP grouping by time, with lines by contour (left) and pitch step (right). Lines by contour overlap, but lines by pitch step show separation.*

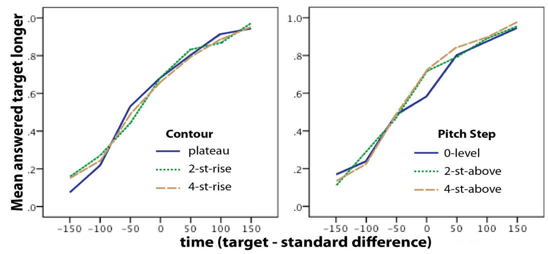

Figure 8: *Experiment 2 results: Mean responses "target longer" by time. Lines by contour (left) overlap. At right, lines by pitch step show some separation at time=0, but otherwise overlap.*

Positive time values indicate a target longer than a standard, and negative that the target is shorter. (For time = 0, target and standard were of identical duration). At left, lines again represent the 3 contours, and at right the 3 pitch steps. The upward diagonal trend of the lines in both graphs shows that subject responses are strongly correlated with the duration, with more responses that the target was longer when the target was indeed objectively longer. The three lines by contour in (figure 8, left) again show virtually no separation, and the 3 lines by pitch step (figure 8, right) only show some suggestion of separation where time = 0.

A mixed model analysis was performed much as in Experiment 1, but with response ("target longer" or "standard longer") as the dependent variable. The result was a model (N = 3360, log-likelihood = -1601) showing significant main effect of time step (target-standard time difference) (Wald Z = 27.47, $p < .0001$). There was no main effect of contour, but also no significant main effect of pitch step. There was, however, a slightly significant interaction between contour and pitch step (Wald Z = 2.43, $p < .05$) only for the 2-st-rise contour with the 4-st-above pitch step.

4. Discussion

The effects of f0 manipulations on grouping perception and duration perception are not identical. The effect of contour (rising pitch vs. plateau) did not strongly influence results in either the grouping perception or the duration perception experiment. Pitch step appears to have influenced responses in both experiments, but potentially in complicated ways.

The complicated interactions between pitch step and contour seen in the results of Experiment 1 are likely a reflection of perceived scaling differences. While pitch step was defined here in terms of pitch distance between the end of *blue* and the following phrase, this likely reflects perceived scaling only for plateau tokens. It is known that for tone glides, perceived pitch is roughly equivalent to a point roughly 2/3 of the way through the glide (Rossi's "2/3 rule," Rossi, 1971 [33]). Estimating perceived pitch thus, tokens where *blue* is either dramatically higher or lower than green will tend to be perceived as B-GP, and where they are similar in pitch there are likely to be more BG-P judgments. Closer pitch across the boundary may influence perceived tonal continuity, and such pitch proximity or similarity of pitch may suggest a weaker boundary. Large pitch changes across a boundary, conversely, create a greater discontinuity, and cue a stronger boundary. Such an observation is in keeping with proposals of prosodic grouping that make reference to gestalt-like principles, such as proposed by Kentner & Féry (2013) [30], and similar to principles proposed for music grouping by Lerdahl & Jackendoff (1983) [34]. It is also worth noting that pitch is recognized as playing a role in connecting phrases into coherent segments in discourses (Wichmann, 2000 [35]; Hansson, 2003 [36]; Hirst, 1993 [37]), but such effects have not been as thoroughly explored in intonational phonology traditions such as AM/ToBI (Beckman & Ayers, 1997 [38]).

In Experiment 2, the lack of effect of pitch contour, seemingly suggests that dynamic pitch does not modulate perceived duration. This may well be due to the orthogonal manipulation of pitch range. When examining the subset of cases where the target and standard durations were equal, more responses that the target was longer can be seen when there is a big pitch jump between target and standard. It is possible that, as shown with the kappa effect, such a pitch jump affects perceived duration of the interceding silent interval. Indeed, it may be the case that subjects are not clearly distinguishing between the durations of the filled intervals vs. the silent intervals. (Note that there is a larger silent interval, 200 *ms* vs. 50 *ms*, in Experiment 2, and this may be reflected in the overall tendency for subjects to hear the first item as longer: 60%, in spite of balanced presentations.) It is possible that at least some of the previously reported effects of dynamic pitch on perceived duration are due to a confound of dynamic pitch introducing pitch step differences across compared tokens. Under certain circumstances dynamic pitch may have an effect, but circumstances may need to be just right, and may be overridden by other effects, such as relative scaling. It is also conceivable that while the psychoacoustic effect of dynamic pitch on perceived duration is real, it may not transfer straightforwardly to speech.

5. Conclusions

These results of these experiments add to our understanding of the effects of pitch on perceived duration, and suggest that pitch factors affect grouping judgments beyond what would be expected from distortions of perceived duration. While pitch factors have been shown to modulate perceived duration, such effects do not account for the degree to which pitch changes affect perceived grouping. Pitch relations across boundaries influence perceived juncture across those boundaries, at times overriding the effects of durational cues. Results support the idea that listeners integrate pitch and timing cues when judging linguistic structure, supporting measures of relative boundary size that combine duration and pitch measures.

6. References

[1] Fon, Y.-J. J., *A Cross-linguistic Study on Syntactic and Discourse Boundary Cues in Spontaneous Speech.* Dissertation. Ohio State University, 2002.

[2] Brown, S. W. "Time and attention: Review of the literature." *Psychology of Time*, 111–138, 2008.

[3] Hoopen, G. T. "Classic Illusions of Auditory Time Perception." *Journal of the Human-Environmental System*, *11*(1), 27–35, 2008.

[4] Lehiste, I.. "Influence of fundamental frequency pattern on the perception of duration." *Journal of Phonetics*, 4, 113–117, 1976.

[5] Cumming, R. "The effect of dynamic fundamental frequency on the perception of duration." *Journal of Phonetics*, 39(3): 375–387, 2011a.

[6] Yu, A., "Tonal effects on perceived vowel duration." In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (Eds.), *Papers in Laboratory Phonology* (Vol. 10). Berlin: Mouton de Gruyter, 2010.

[7] Henry, M. *A Test of an Auditory Motion Hypothesis for Continuous and Discrete Sounds Moving in Pitch Space*. Dissertation. Bowling Green State University, 2011.

[8] Cohen, J., Hansel, C., & Sylvester, J. "A new phenomenon in time judgment", *Nature*, *172*: p. 901, 1953.

[9] Cohen, J., Hansel, C. & Sylvester, J. "Interdependence of temporal and auditory judgments." *Nature*, *174*: 642–644, 1954.

[10] Shigeno, S. "The interdependence of pitch and temporal judgments by absolute pitch possessors." *Perception & Psychophysics*, 54(5): 682–692, 1993.

[11] Crowder, R. & Neath, I. "The influence of pitch on time perception in short melodies", *Music Perception*, 12(4): 379–386, 1995.

[12] MacKenzie, N. *The kappa effect in pitch/time context*. Dissertation. Ohio State University, 2007.

[13] Brugos, A. & Barnes, J. "The auditory kappa effect in a speech context." Speech Prosody, Shanghai, China. 2012b.

[14] Brugos, A. & Barnes, J. "Pitch trumps duration in a grouping perception task," 25th Annual CUNY Conference on Human Sentence Processing, New York, NY. 2012a.

[15] Jun, S. "Prosodic phrasing and attachment preferences." *Journal of Psycholinguistic Research*, *32*(2), 219–249, 2003.

[16] Beach, C. M. "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, 1991.

[17] Cumming, R. E. "The interdependence of tonal and durational cues in the perception of rhythmic groups," *Phonetica* 67, 219–242, 2011b.

[18] Jeon, H.-S., & Nolan, F.. "The role of pitch and timing cues in the perception of phrasal grouping in Seoul Korean." *The Journal of the Acoustical Society of America*, *133*(5), 3039-3049, 2013.

[19] Scott, D. R. "Duration as a cue to the perception of a phrase boundary." *Journal of the Acoustical Society of America*, *71*(4), 996–1007, 1982.

[20] Wagner, M., & Crivellaro, S. "Relative Prosodic Boundary Strength and Prior Bias in Disambiguation." in *Proceedings of Speech Prosody 2010*. Chicago, 2010.

[21] House, D., *Tonal Perception in Speech.* Lund, Sweden: Lund University Press, 1990.

[22] Jeon, H.-S., and Nolan, F. "Segmentation of the Accentual Phrase in Seoul Korean," in *Proceedings of Speech Prosody 2010*, Vol. 100023, pp. 1–4, 2010.

[23] Beach, C. M., Katz, W. F., & Skowronski, A. "Children's processing of prosodic cues for phrasal interpretation." *Journal of the Acoustical Society of America*, *99*(2), 1148–1160, 1996.

[24] Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. "The use of prosody in syntactic disambiguation." *Journal of the Acoustical Society of America*, 90, 2956-2970, 1991.

[25] Clifton, C. J., Carlson, K. and Frazier, L., "Informative prosodic boundaries," *Language and Speech*, vol. 45, pp. 87–114, 2002.

[26] Selkirk, E. "On derived domains in sentence phonology." *Phonology Yearbook*, 3: 371-405, 2986

[27] Ladd, D. R. "Intonational phrasing: The case for recursive prosodic structure." *Phonology* 3. 311–340, 1986.

[28] Wagner, M. *Prosody and recursion*. Dissertation, MIT, 2005.

[29] Schreuder, M. J. *Prosodic processes in language and music*, Dissertation, Groningen University, 2006.

[30] Kentner, G., & Féry, C. "A new approach to prosodic grouping." *The Linguistic Review*, *30* (2), 2013.

[31] Bates, D. & Maechler, M. "lme4: Linear mixed-effects models using S4 classes." R package version 0.999375-32, 2009.

[32] Baayen, R., Davidson, D. & Bates, D. "Mixed-effects modeling with crossed random effects for subjects and items", *Journal of Memory and Language*, 59: 390-412, 2008.

[33] Rossi, M. "Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole," *Phonetica*, **23**, 1-33, 1971.

[34] Lerdahl, F., & Jackendoff, R. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press, 1983.

[35] Wichmann, A. *Intonation in Text and Discourse*: *Beginnings, Middles and Ends.* Harlow, UK: Longman, 2000.

[36] Hansson, P., *Prosodic Phrasing in Spontaneous Swedish*. Dissertation, Lund University, 2003.

[37] Hirst, D. "Peak, boundary and cohesion characteristics of prosodic grouping." *ESCA Workshop on Prosody*, 1993.

[38] Beckman, M., & Ayers Elam, G. *Guidelines for ToBI Labelling* (version 3, March 1997).