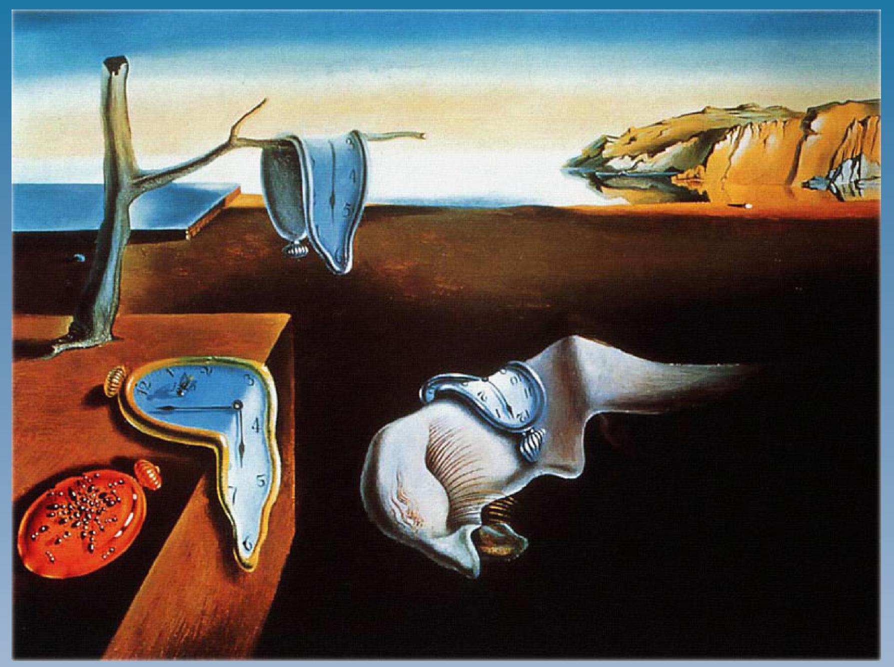
Effects of dynamic pitch and relative scaling on the perception of duration and prosodic grouping in American English

Alejna Brugos **Jonathan Barnes**
Boston University, Boston, Massachusetts, USA



Introduction

Duration measures have been key to the study of prosodic boundary size:
- Increased pre-boundary lengthening and pause duration correlated with stronger prosodic boundaries in production & perception (Wightman et al 1992; Lehiste et al 1996; inter alia)

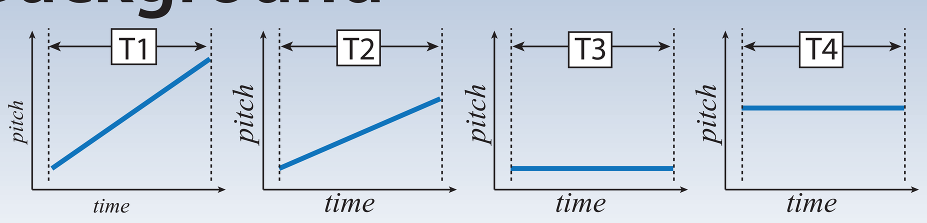
Pitch features also mark prosodic domain edges (phrase tones and reset)

Phonetic measures of pitch and timing are typically taken independently
However, pitch has been shown previously to affect perceived duration:
- Tokens with dynamic f0 heard as longer than those with static f0
- Relative pitch also known to distort perceived duration

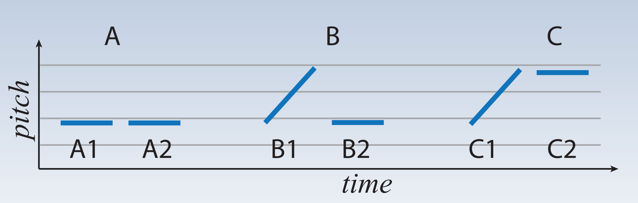
As prosodic boundaries are often marked by dynamic f0 (phrase tones)
- How might the effects of such pitch dynamicity interact with preboundary lengthening in perception of duration?
- How might the same manipulations affect linguistic judgments of prosodic grouping?

Two new experiments used the same duration and f0 manipulations of segmentally-identical base files, in two separate tasks:
1) a linguistic grouping task using an ambiguously-structured phrase
2) a psychoacoustic study on perceived duration.

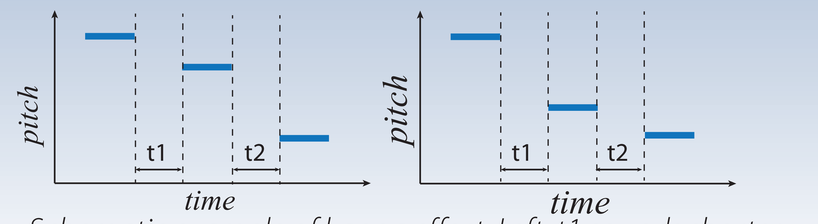
Background



Dynamic pitch and scaling affect perceived duration:
objective time: T1=T2=T3=T4 *perceived time:* T1>T2>T3; T4>T3



Dynamic pitch changes pitch relations across tokens



Schematic example of kappa effect: Left, t1 sounds shorter than t2; Right, t1 sounds longer than t2.

Perception of time can be systematically affected by a range of contextual factors (Brown, 2008), including pitch (Hoopen, 2008)

1. **Dynamic pitch and scaling differences affect perceived duration of filled intervals**
   - Vowels with dynamic pitch perceived as longer than those with static pitch (Lehiste, 1976; Yu, 2010; Cumming, 2011; inter alia)
   - Vowels with higher pitch perceived as longer than those with lower pitch (Yu, 2010)
   - Non-speech tone glides with greater pitch change velocity heard as longer than than tone glides with lesser pitch change velocity (Henry, 2011)
2. **Comparing dynamic pitch to level pitch intervals introduces differences in pitch across tokens**
3. **Pitch differences between tokens across silence distorts perception of silent intervals (the auditory kappa effect):**
   - Silent intervals bounded by tones of closer pitch are perceived as shorter than those bounded by those of more widely differing pitch in both speech (Brugos & Barnes, 2012a) and nonspeech contexts (Cohen et al,1953; Shigeno,1993; MacKenzie, 2007; inter alia)

Further, pitch relations across silent intervals can affect perceived prosodic grouping beyond their effects on duration perception (Brugos & Barnes, 2012b)

Question: How to overcome confound to explore effects of phrase-final dynamic pitch on perceived duration and grouping?
Answer: Manipulate relative scaling along with dynamicity directly.

Method: Two Matched Perception Experiments

Experiment 1: linguistic judgment of grouping
- Using ambiguous phrase: *blue and green and purple*
- Can be parsed as *blue and (green and purple)* (B-GP) or *(blue and green) and purple* (BG-P)

Expections based on relative perceived duration:
➔ blue > green, more B-GP judgements,
➔ blue < green, more BG-P judgements

Will pitch proximity play a role as it did in Brugos & Barnes 2012b?

Experiment 2: a psychoacoustic judgment of perceived duration

Expection: Dynamic pitch tokens will sound longer than those with static pitch, and effect on perceived duration will increase with degree of dynamicity.

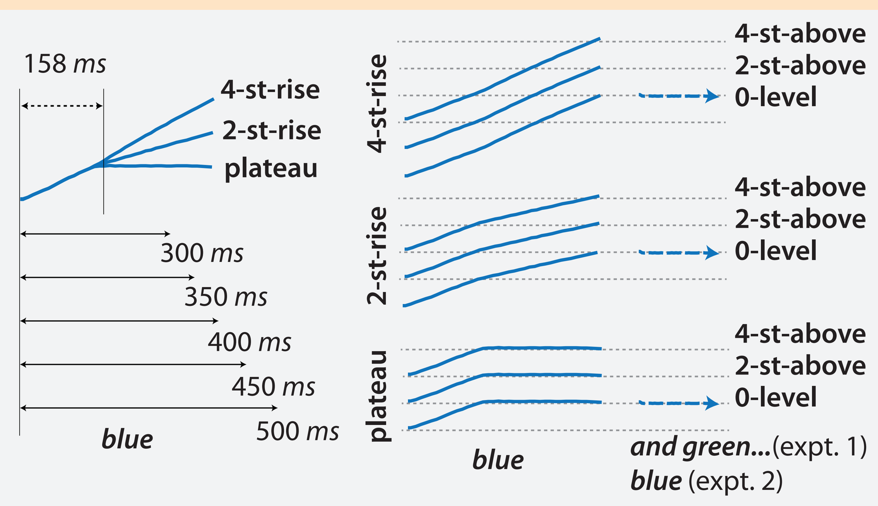
How will known effects of relative pitch scaling interact?

Experiment 1:

Stimuli: resynthesized versions of same base file of phrase *blue and green and purple*

Target: *blue* with crossed continua of duration and f0 manipulations (see image and description at right)

Presentation & Task:
- Targets concatenated with phrase completion *and green and purple* with fixed f0 pattern (see figure)
- 3 durations of word *green* (~350, ~400, ~450)
- 360 trials, randomized for each subject
- 16 subjects responded via button box
- Asked "which grouping?" (illustrated via images.)



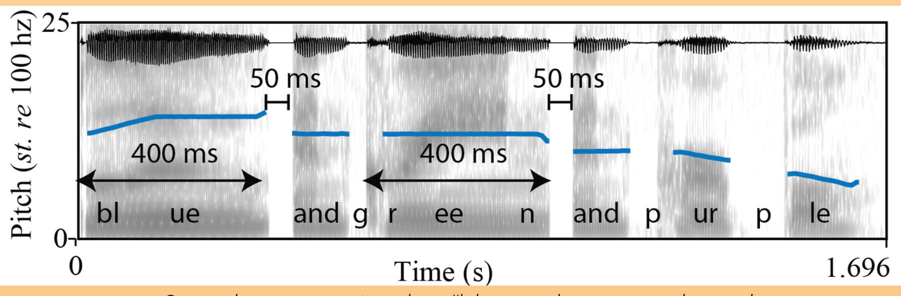
Target for both experiments: resynthesized versions of spoken word *blue* uing same base file:
- 3 crossed continua of manipulations:
  ➔ 5 durations (300, 350, 400, 450, 500 ms)
  ➔ 3 f0 contours (plateau and 2 rises)
  ➔ 3 pitch ranges for each contour

Experiment 2:

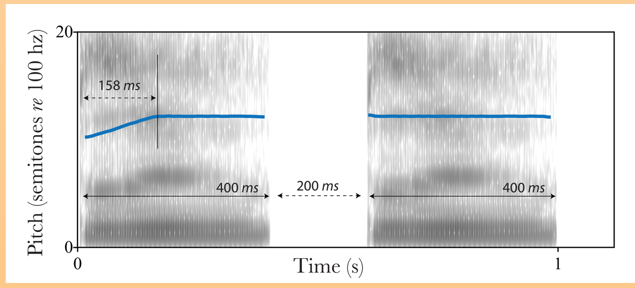
Stimuli: pairs of manipulations of spoken word *blue*

Target: identical to manipulations of *blue* from expt. 1
Standard: flat f0 standards at 5 durations, at same f0 level as *green* in expt. 1

Presentation & Task:
- Presented in 2 orders: target first and standard first
- 200 ms silence between
- 600 trials, randomized for each subject
- 9 subjects (subset of subjects from expt. 1)
- responded via laptop keys
- Asked of presented pair: "Which sounds longer?"



Sample expt. 1 stimulus: "blue and green and purple



Sample expt. 2 stimulus: "blue, blue"

"Which sounds longer?"

Results & Analysis

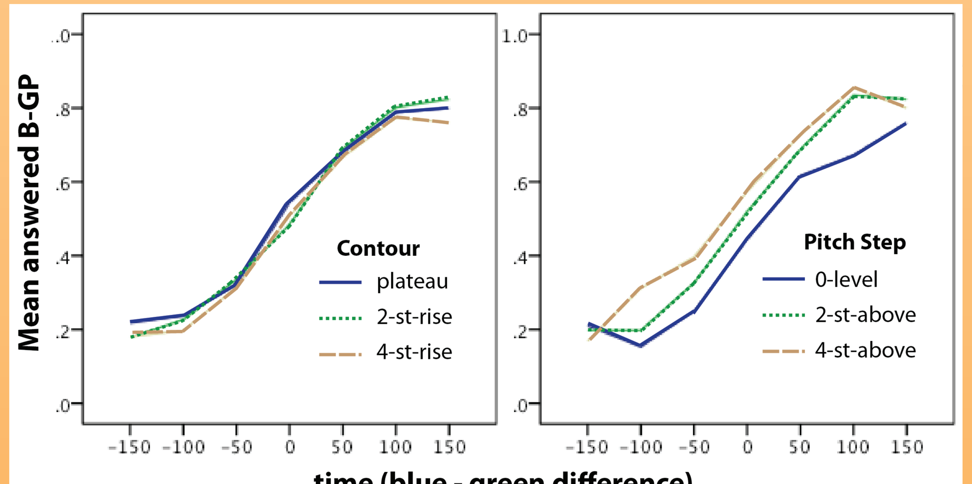
Experiment 1: results from 5598 trials (N=16)
Experiment 2: results from 3360 trials (only target-standard order shown) (N=9)

Results graphed analogously for Experiment 1 (left) and Experiment 2 (right)

X axis: Time difference between *blue* and *green* (expt. 1) or target and standard *blue* (expt. 2): Positive time values indicate that target *blue* is longer than *green* (expt. 1) or standard *blue* (expt. 2), and negative that target *blue* is shorter.

Y axis: Mean responses B-GP (expt.1) or "target longer" (expt. 2)

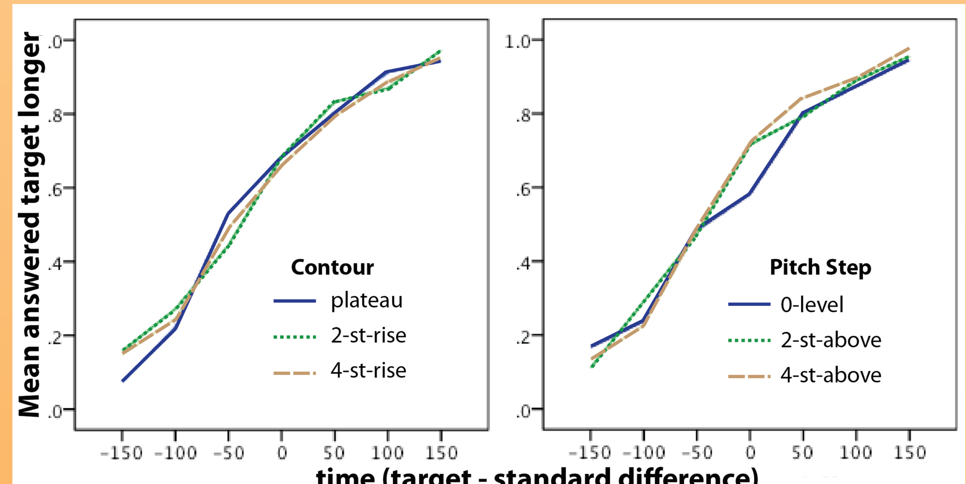
Experiment 1:



Upward diagonal trend of lines shows that responses are strongly correlated with duration difference between target and green/standard.

Experiment 1 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.

Experiment 2:



Note: an expanded version of Experiment 2 with additional subjects (N=20) and one additional pitch step (2 st below) did show separation by pitch step to a significant level, though still no measurable effect of contour (Brugos & Barnes, 2014)

Experiment 2: 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.

Discussion & Conclusions

Results from both experiments together suggest that listeners may be responding to perceived prosodic distance that integrates cues from timing (filled and silent intervals) and pitch (pitch slope and pitch jumps across silent intervals)

- Previously demonstrated effects on perceived duration due to dynamic pitch not straightforwardly reproduced:
  ➔ perceived duration of tokens potentially more affected by relative scaling of compared tokens (cf. Brugos & Barnes, 2014)
- Same manipulations push grouping judgments beyond what would be expected from distortions of perceived duration.
  ➔ Pitch relations across boundaries influence perceived juncture across boundaries: phrases closer in pitch tend to be grouped together
  ➔ At times even overriding the effects of durational cues

Results suggestive of cue trading relations in pitch and timing boundary cues (Beach, 1991, Jeon & Nolan, 2013)

Results also reminiscent of recognized role that pitch plays connecting phrases into coherent segments in discourses (Wichmann, 2000; Hansson, 2003 ; Hirst, 1993)

Results are compatible with approaches to prosodic grouping that make reference to gestalt-like principles (eg. proximity, similarity, continuity) (Jeon & Nolan, 2013), such as proposed by Kentner & Féry (2013), and similar to principles proposed for music grouping by Lerdahl & Jackendoff (1983)

Listeners integrate pitch and timing cues when judging linguistic structure, supporting measures of relative boundary size that combine duration and pitch measures.

Measures of prosodic boundary strength should not rely exclusively on objective duration.

References

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.
Brown, S. W. "Time and attention: Review of the literature." *Psychology of Time*, 111–138, 2008.
Brugos, A. & Barnes, J. "Dynamic pitch and pitch range interact in distortions of perceived duration of American English speech tokens," ASA, Providence, May, 2014.
Brugos, A. & Barnes, J. "The auditory kappa effect in a speech context." Speech Prosody, Shanghai, China. 2012a.
Brugos, A. & Barnes, J. "Pitch trumps duration in a grouping perception task," CUNY 25, NY. March, 2012b.
Cohen, J., Hansel, C., & Sylvester, J. "A new phenomenon in time judgment", *Nature*, 172: p. 901, 1953.
Cumming, R. "The effect of dynamic [f0] on the perception of duration." *JPhon*, 39(3): 375–387, 2011.
Henry, M. *A Test of an Auditory Motion Hypothesis for Continuous and Discrete Sounds Moving in Pitch Space*. Dissertation. Bowling Green State University, 2011.
Hansson, P., *Prosodic Phrasing in Spontaneous Swedish.* Dissertation, Lund University, 2003.
Hirst, D. "Peak, boundary and cohesion characteristics of prosodic grouping." ESCA Workshop on Prosody, 1993.
Hoopen, G. T. "Classic Illusions of Auditory Time Perception." *Journal of the Human-Environmental System*, 11(1), 27–35, 2008.
Jeon, H.-S., & Nolan, F.. "The role of pitch and timing cues in the perception of phrasal grouping in Seoul Korean." *JASA*, 133(5), 3039-3049, 2013.
Lehiste, I.. "Influence of fundamental frequency pattern on the perception of duration." *JPhon*, 4, 113–117, 1976.
Lehiste, I., Olive, J. P., & Streeter, L. A. "Role of duration in disambiguating syntactically ambiguous sentences." *JASA*, 60(5), 1–4, 1976.
Lerdahl, F., & Jackendoff, R. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press, 1983.
MacKenzie, N. *The kappa effect in pitch/time context. Dissertation*. Ohio State University, 2007.
Shigeno, S. "The interdependence of pitch and temporal judgments by absolute pitch possessors." *Perception & Psychophysics*, 54(5): 682–692, 1993.
Wichmann, A. *Intonation in Text and Discourse: Beginnings, Middles and Ends*. Harlow, UK: Longman, 2000.
Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. "Segmental durations in the vicinity of prosodic phrase boundaries." *JASA*, 91: 1707–1717, 1992.
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.