Perception of prosodic boundaries in spontaneous speech with and without silent pauses

Yoonsook Mo, Jennifer Cole
Beckman Institute, Department of Linguistics, University of Illinois at Urbana-Champaign

Given the multiplicity of acoustic cues to prosody,

Q1) Is the perception of prosodic boundary dependent on silent pause as an independent cue?
Or
Q2) Does redundancy in the acoustic encoding of prosody support robust prosody perception in the absence of silent pauses?

Introduction

- In everyday conversation, speakers communicate pragmatic and discourse meaning through the prosodic form assigned to an utterance.
- Prosody is encoded in the acoustic signal in suprasegmental properties and also influences the acoustic realization of segmental features within words.
- Listeners must attend to the acoustic cues to prosodic form to fully recover the speaker's intended meaning.

Goals of this study:

1. to measure the contribution of the information in the acoustic signal to the perception of prosodic boundaries in spontaneous conversational speech.
2. to evaluate whether the presence of silent pause is necessary to perceive prosodic boundaries in real-time prosody perception.

- Many prior studies employing controlled "laboratory" speech (e.g. simple sentences, read speech) show that silent pause is a primary correlate of prosodic boundary and important for distinguishing levels of prosodic boundary.
- Other studies employing spontaneous conversational speech show that speakers vary the location and the length of silent pauses and listeners can reliably detect prosodic boundaries even without pause.
- No study has investigated the cues to boundary perception in online, real-time processing of spontaneous speech.

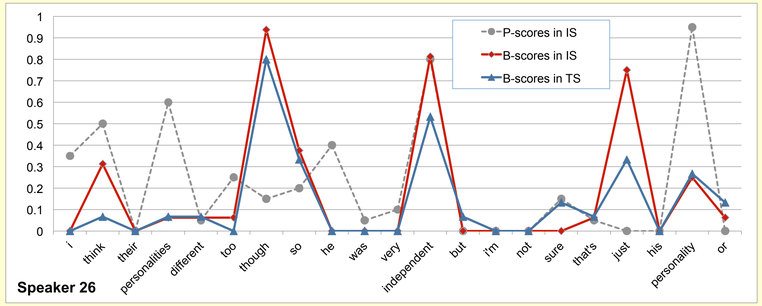
Accepting that silent pauses are a primary cue for boundary perception, do untrained nonexpert listeners rely on pause information, or can they reliably detect prosodic boundaries based on acoustic information other than silent pauses?

Rapid Prosody Transcription (RPT)

1. 36 short excerpts (~11 – 23 sec) were selected from the Buckeye corpus of American English spontaneous speech: 2 excerpts *19 speakers
   Intact Stimuli (IS): Speech excerpts as originally produced
   Pause-Truncated Stimuli (PTS): Speech excerpts where silent pauses longer than 20ms are removed
2. Orthographic transcripts were produced for each sound file, with no punctuation or capitalization.
3. Subjects were UIUC undergraduates, untrained and unfamiliar with the phonetics and phonology of prosody. 74 subjects in three groups of 14-16 were in the IS condition, and one group of 15 subjects in the TS condition.
4. After being given simple instructions and definitions of prominence and boundary, subjects marked the locations of prosodic boundaries on the printed transcripts in real time, based only on auditory impression (no visual speech display).
5. Using the transcription data from all subjects, each word in the IS and TS excerpts was assigned probabilistic P(rominence) - and B(oundary) -scores depending on the number of transcribers who marked a word as prominent or as followed by a juncture.

Results

Comparisons of P- and B-scores in IS vs. B-scores in TS



Comparisons of Reliability tests

z=2.32, α=0.01		Intact Stimuli				Truncated Stimuli
		Session 1	Session 2	Session 3	Session 4	
Boundary	Kappa	0.587	0.609	0.532	0.640	0.374
	z	**31.3**	**33.0**	**26.8**	**37.3**	**25.5**

1. Fleiss' multi-rater's kappa agreement scores in the IS condition are much above chance level.
   → *Untrained, nonexpert listeners' perception of boundary is systematic and reliable.*
2. Kappa scores in the TS condition are also much above chance level.
   → *Untrained, nonexpert listeners can reliably and systematically locate boundaries even without silent pauses from the speech.*
3. Average kappa scores in the IS condition are always considerably higher than the kappa score in TS.
   → *Silent pause is an important cue for boundary perception.*
   → *Yet untrained, nonexpert listeners' perception of boundary is also cued by acoustic parameters other than silent pause in spontaneous conversational speech of American English.*

Boundary detection on each word: IS vs. TS

		Truncated Stimuli		Total
		Boundary	Non-boundary	
Intact Stimuli	Boundary	123	747	870
	Non-boundary	307	56	363
Total		430	803	1233

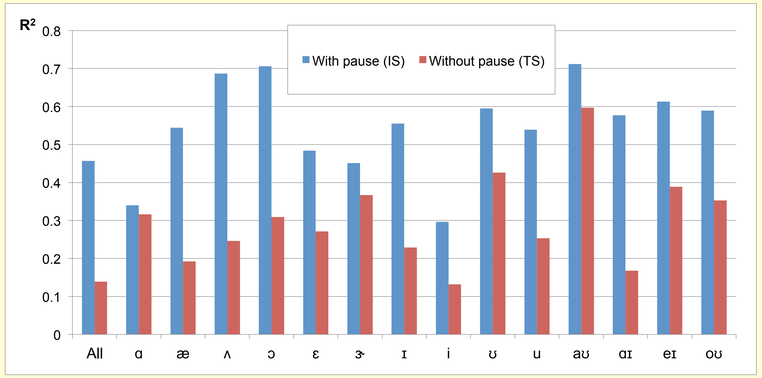
Cohen's kappa agreement score on the detection of prosodic boundary with and without silent pauses is **0.668.**

→ Untrained, nonexpert listeners perceived prosodic boundaries in TS in the same locations where prosodic boundaries are perceived in IS.

Acknowledgements
This research is supported by NSF grants IIS 07-03624 to Jennifer Cole and Mark Hasegawa-Johnson. Special thanks to the Prosody-ASR group members for their comments

Regression models of perceived prosodic boundaries as predicted by acoustic measures

- Acoustic measures are extracted from the word final stressed vowels.
- Stress indentified according to the ISLE dictionary(Hasegawa-Johnson and Fleck, 2007)
- Normalized acoustic measures
  - vowel duration, overall intensity, local F0 maximum, F0 at the right edge
  - sub-band intensities (0-0.5, 0.5-1.0, 1.0-2.0, 2.0-4.0 kHz)



- Prosodic boundaries as identified by untrained nonexpert listeners both in IS and in TS are modeled on the basis of acoustic information available in the speech signal.
- Between 30 -71% of the variability in boundary perception in IS is explained by acoustic measures including silent pauses.
- Between13 - 60% of the variability in boundary perceptoin in TS is explained by the acoustic measures excluding silent pauses.
- Between 30 - 93% of the variability explained by the regression models in IS is explained by the regression models in TS.

Discussion and conclusion

1. In real-time online prosody perception with spontaneous conversational speech, untrained nonexpert listeners can detect the location of prosodic boundaries, based on the acoustic information including silent pauses.
2. Given the multiplicity of acoustic cues, untrained nonexpert listeners can also identify the locations of prosodic boundaries with no silent pauses.
3. Although listeners can perceive prosodic boundaries without silent pauses, the existence of silent pauses in spontaneous conversational speech aids them to more reliably identify the locations of prosodic boundaries.

Silent pause is an important cue for the perception of prosodic boundaries, but redundancy in the acoustic encoding of prosody supports boundary perception in the absence of the primary cue.

References
1. Choi, J-Y., 2003. Pause duration and speech rate as durational cues for prosody markers. *JASA*, 114 (4), 2395-2395.
2. Cho, H. and Hirst, D. 2006. The contribution of silent pauses to the perception of prosodic boundaries in Korean read speech. In proceedigs of Speech Prosody (Dresden, Germany), 185-188.
3. Collier, R., Pijper J. R. de, and Sanderman, A., Perceived prosodic boundaries and their phonetic correlates. In proceedings of the workshop on human language technology (Princeton, NJ), 341-345.
4. Hasegawa-Johnson, M.; Fleck, M., ISLE Dictionary version 0.2.0," 2007, downloaded Oct. 19, 2007 from http://www.isle.uiuc.edu/dict/index.html
5. Li, W. and Yang, Y., 2009. Perception of prosodic hierarchical boundaries in Mandarin Chinese setences, Neurosience, 158 (4), 1416-1425.
6. Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E., 2007. *Buckeye Corpus of Conversational Speech* (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).
7. Tseng, C-Y. and Chang, C-H., 2008. Pause or no pause? –Prosodic phrase boundaries revisited. *Tsinghua Science and Technology*, 13 (4), 500-509.
8. Viola, I. C. and Madureira, S., 2008. The roles of pause in speech expression. In proceedings of Speech Prosody (Campinas, Brazil), 721-724.
9. Yoon, T-J., Cole, J, and Hasegawa-Johnson, M., 2007. On the edge: Acoustic cues to layered prosodic domains. In proceedings of ICPhS (Saarbruken, Germany), 1017-1020.