How do ordinary listeners perceive prosodic prominence? Syntagmatic vs. Paradigmatic comparison

Yoonsook Mo[1], Jennifer Cole[1], Mark Hasegawa-Johnson[2]

[1]Department of Linguistics, University of Illinois at Urbana-Champaign

[2]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign

Q1) What acoustic feature or feature combinations cue prosodic prominence?

Q2) Which normalization method do ordinary listeners adopt for prominence perception?

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:

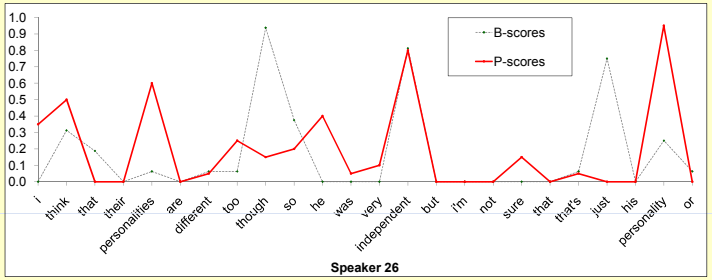
1. to identify the acoustic cues for prosodic prominence, and their individual and joint contributions to prominence perception
2. to examine whether or how acoustic cues are normalized to control for variation due to speaker and local context in ordinary listeners' perception of prosodic prominence

- Most prior studies employing controlled "laboratory" speech (e.g. simple sentences, read speech) show acoustic effects of prosody in many languages, including effects of prosodic prominence, and also show that native listeners respond to acoustic prosodic cues in interpreting utterance meaning (See references below, among many others)
- Very few studies evaluate how acoustic variation due to other sources (e.g. speaker, or local phonological, syntactic, or discourse context) interacts with prosody perception.

Do listeners perceive prosodic cues relative to the local context, or relative to the individual speaker's baseline, or based only on raw cue vales?

Methodology

1. 54 short excerpts (~11 – 58 sec) were selected from the Buckeye corpus of American English spontaneous speech (Pitt et al., 2007).
2. 97 undergraduates at UIUC marked the locations of prominences and boundaries while listening to speech excerpts based only on auditory impression in real time.
3. After transcription tasks, each word 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.



Acknowledgements

This research is supported by NSF grants IIS 07-03624 and IIS 04-14117 to Jennifer Cole and Mark Hasegawa-Johnson. Special thanks to Eun-Kyung Lee for data collection and Prosody-ASR group members for their comments

Reliability test

z=2.32, α=0.01		Exp.1		Exp. 2		Exp. 2	
		Grp.1	Grp.2	Grp.3	Grp.4	Grp.5	Grp.6
Prominence	Kappa	0.373	0.421	0.394	0.407	.356	.400
	z	**19.43**	**20.48**	**18.15**	**18.31**	**15.31**	**19.56**

Fleiss' multi-rater's kappa agreement scores are much above a chance level: *ordinary, untrained listeners' perception of prominence is **systematic and reliable**.*

Acoustic measurements

- Acoustic measures are extracted from stressed vowels to hold stress information constant
- Stress indentified according to the ISLE dictionary. (Hasegawa-Johnson and Fleck, 2007)
- List of acoustic measures
  - duration, overall intensity, F0 maximum
  - sub-band intensities (0-0.5, 0.5-1.0, 1.0-2.0, 2.0-4.0 kHz)

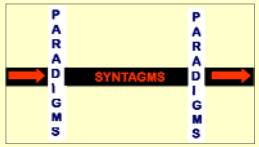
Normalization

1. Paradigmatic normalization

by phone identity within speaker

z-normalization $z = \frac{x - \bar{x}}{s}$

γ-normalization $g = \frac{x}{\bar{x}}$



- Raw acoustic measures and logarithmic transforms are z-normalized.
- Raw acoustic measures are also γ-normalized.

2. Syntagmatic normalization

Raw acoustic measures are **z-normalized** in the **local context within an utterance.**

Target vowel is located at the center.

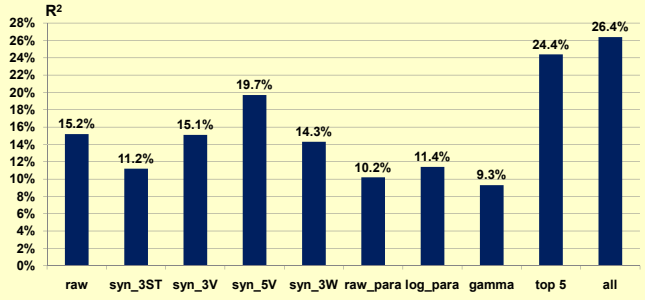
Domain size: 3 and 5 vowels, 3 words, and 3 stressed vowels

Statistical analyses

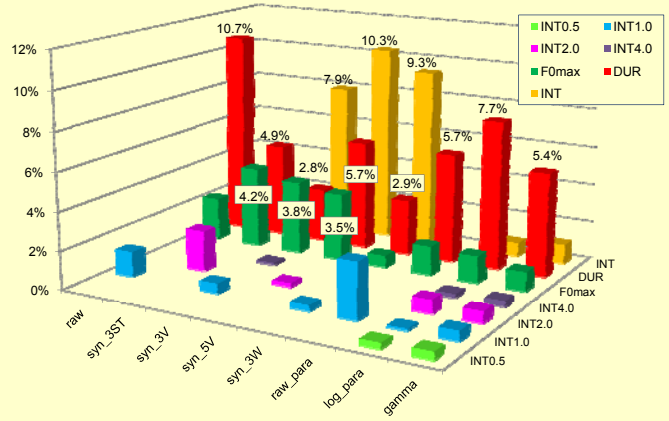
Simple and stepwise multiple linear regressions of the acoustic measures as P-scores

Results

Total variation in P-scores accounted for based on acoustic measures



Contribution of each acoustic measure to prominence perception



1. The acoustic measures normalized syntagmatically in a domain of 5 neighboring vowels account for the largest variation (19.7%) of listeners' perception of prominence compared to any other normalization method.
2. The acoustic measures that are not normalized at all account for 15.2% of variation in perceived prominence by listeners.
3. The 5 top most important cues to prominence can predict 24.4% of variation in listeners' response to prominence; increasing the number of predictors in the regression models does not help much (26.4%).
4. Raw duration accounts for the largest variation (10.7%) of listeners' perception of prominence as a single cue; overall intensity shows up as a major cue for prominence only when syntagmatically normalized (syn_3V: 7.9%, syn_5V: 10.3%, syn_3W: 9.3%)

Discussion and conclusion

1. Untrained ordinary listeners do perceive prosodic prominence systematically and reliably.
2. When listeners perceive prominence, they may or may not employ normalization methods in the processing of the acoustic cues depending on the characteristics of each acoustic cue; listeners must be attentive to raw durations as well as changes in loudness in a local context within a utterance.
3. The 5 top most important cues to prominence can account for a sizable portion of the total variation of listeners' perception of prominence (24.4%).
4. Perceived prominence can be predicted by combinations of acoustic measures but there is not a single acoustic measure that emerges as the primary cue.
5. In perceiving prominence, listeners attend to local changes in overall intensity and F0 max, but duration is a robust cue to prosody, with or without contextual normalization.

References

1. Beckman, M. E., 1986. *Stress and non-stress.* Dordrecht, The Netherlands: Foris Publications.
2. Beckman, M. E., Edwards, J., & Fletcher, J., P1992. rosodic structure and tempo in a sonority model of articulatory dynamics. In Doherty, G. J. and Ladd, D. R. (eds). *Papers in Laboratory Phonology*II: *Gesture, segment, Prosody*, 68-86, Cambridge University Press, Cambridge
3. Cho, T., 2005. Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *JASA*, 117 (2), 3867-3878
4. Fry, D. B., 1958. Experiments in the perception of stress. Language and speech. 1. 126-152.
5. Gusssenhoven, C. and Rietveld, A. C. M, 1988. Fundamental frequency declination in Dutch: testing three hypotheses. *J. Phonetics*. 16, 355-369.
6. Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H. & Terken J. 1997. The perceptual prominence of fundamental frequency peaks, *JASA*, 102 (5). 3009-3022
7. 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
8. Heldner, M., 2003. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *J. Phonetics*. 31, 39-62.
9. Heldner, M. and Strangert, E., 1997. TO what extent i perceived focus determined by F0-cues? Proceedings of Eurospeech (Rhodes, Greece), 2, 875-877.
10. Hermes, D. J. and Rump, H. H., 1994. Perception of prominence in speech intonation induced by rising and falling pitch movements. *JASA*, 90, 38-44.
11. 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).
12. Sluijter, A. M. C. and van Heuven, V. J., 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of ICSLP* (Philadelpia, PA). 630-633.
13. Turk, A. E.; Sawusch, W., 1996. The processing of duration and intensity cues to prominence. *JASA*, 99 (6), 3782-3790.