Optimal models of prosodic prominence using the Bayesian information criterion

Tim Mahrt, Jui-Ting Huang, Yoonsook Mo, Margaret Fleck,
Mark Hasegawa-Johnson, and Jennifer Cole
Beckman Institute, University of Illinois

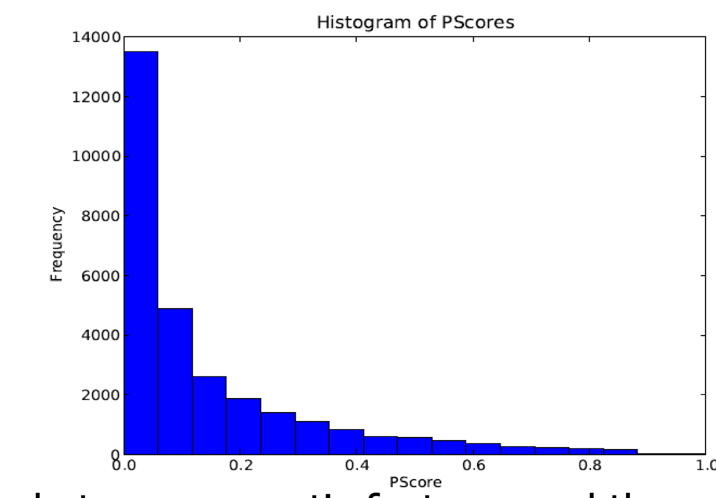
Beckman Institute

Background:

- Past research has suggested that duration, pitch, and intensity, along with speaker expectations, all play a role in the perception of prominence [1,2,3,4,5,6,9]. The goal of this study is to understand how variation along the continuous parameters related to these factors relates to the variable perceptual responses of listeners. Specifically, we ask if these factors signal gradient or discrete distinctions in perceived prominence. This question has implications for the phonological encoding of prominence.

Data:

- In our previous work, we collected prominence judgments for a 35,009 word subset of the Buckeye Corpus of spontaneous speech [7], using Rapid Prosody Transcription, a method we developed for prosodic labeling of spontaneous speech [8].
- Excerpts were transcribed for prosodic prominence by teams of 15-20 naive speakers of English. Their task was to label each word as prominent or non-prominent in real time as they listened to short (15-60 s) excerpts.
- For each word, we added up the number of labelers who labeled the word as prominent and then divided this sum by the number of labelers for that word. We call this value the p-score.
- The distribution of p-scores over words in the database:



Previous Findings:

- In a later study, we looked for a relation between acoustic features and the ratings of prominence and found a positive correlation between them; as acoustic values become more extreme the number of listeners who rate the word as prominent increases as well [5].
- The results of that study show that p-scores (i.e., measures of the likelihood that a word will be perceived as prominent) co-vary with word frequency and with the acoustic cues to prominence.

Models of Prominence:

- Binary Model: Two distributions--prominent and non-prominent. In collecting prominence ratings from subjects, we assumed a binary model.
- Continuous Model: A single distribution over a continuous valued prominence feature. The assignment of p-scores assumes a quasi-continuous model.

Features for this Study:

- Word Duration, Stressed Vowel Duration, Word Frequency, Post-Word Pause Duration, and Pre-Word Pause Duration

Bayesian Information Criterion:

- To determine which model of the distribution is optimal for each feature we perform a series of model comparisons. For models with the same number of distributions, a likelihood function is used to find the better model.
- However, to compare models containing differing number of distributions, the Bayesian Information Criterion (BIC) is used. The BIC is a likelihood estimate with a penalty for having a larger number of mixtures [10].

$$BIC(X;\Lambda) = logF(X;\Lambda) - (k/2)ln(n)$$

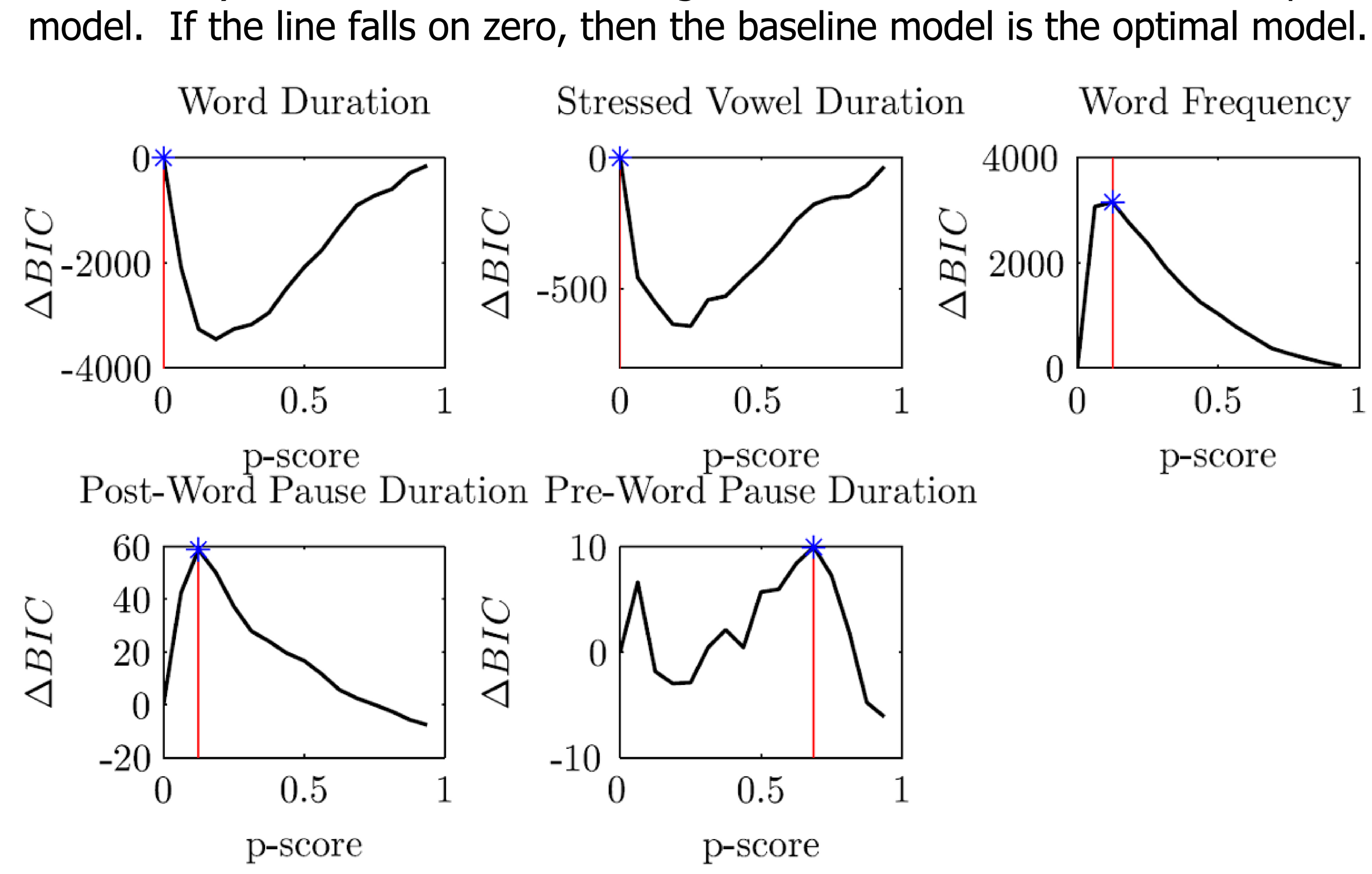
- To compare two models we compute

$$\Delta BIC = BIC(X;\Lambda_1) - BIC(X;\Lambda_2)$$

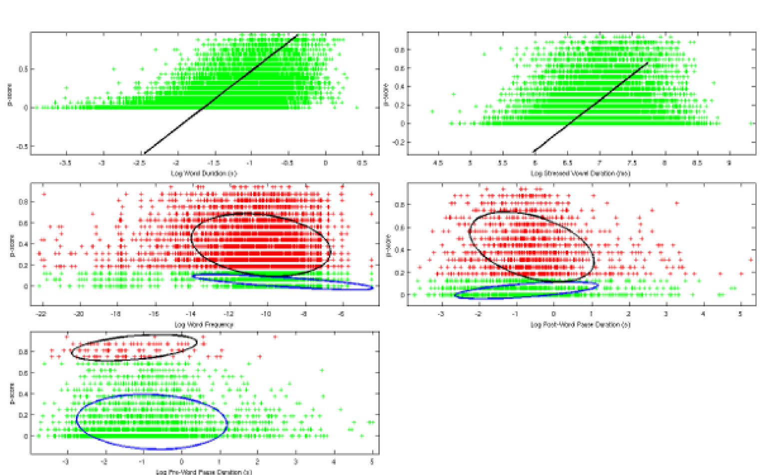
- If model 1 is better than model 2, than delta BIC will be positive.

Results:

- The following graphs show the results for the features analyzed. P-score thresholds, which distinguish models, are plotted against BIC scores. BIC scores are reported compared to the baseline model (the model with only a single distribution).The red line lies at the highest BIC score and indicates the optimal model. If the line falls on zero, then the baseline model is the optimal model.



The following shows scatter plots of the Features plotted against p-values. Each plot has a Gaussian ellipse over the distribution. All feature values, except for word frequency are log values of the original feature values.



Optimal Models of Prominence

Feature	Model	P-score threshold
Word duration	Single Gaussian	N/A
Stressed vowel duration	Single Gaussian	N/A
Word frequency	Two Gaussians	0.125
Post-word pause duration	Two Gaussians	0.125
Pre-word pause duration	Two Gaussians	0.6875

Conclusions:

- The results show that some features known to be correlates of prominence are better modeled by two Gaussian distributions rather than one, that features have very different BIC score distributions, and that the optimal point to partition a feature varies depending on the feature.
- In order to account for these differences one of the following must be true:
  - Listeners differ in the meaning-based or acoustic criteria they employ to identify infer prominence. This is supported by our previous work [5]. In that study, we predicted p-scores from a set of acoustic measures and found that the contribution of each feature to the variability in p-scores varies from speaker to speaker.
  - Prominence may not be binary but could be n-ary. This would suggest that there are different kinds of acoustic correlates to different kinds of prominence (e.g. prominence due to new information or contrastive focus).
  - In part, listeners differ in the cues that they use and in part, prominence is n-ary. In our view, this is the most likely. If the perception of prominence is purely idiosyncratic then prominence would not be useful for carrying linguistic information but there is already evidence that to at least some degree it is idiosyncratic [5].
- All features have a local maximum at a low p-score. This local maximum aligns with our prediction that there is a meaningful distinction between low-valued and high-valued p-scores.

References

[1] Kochanski, G., Grabe, E., Coleman, J. and Rosner., B., “Loudness predicts prominence: Fundamental frequency lends little”, The Journal of the Acoustical Society of America, vol 118, 2005.
[2] Ladd, D.R., “Intonational phonology”, Cambridge University Press, 2008.
[3] Calhoun, S., “Information structure and the prosodic structure of English”, University of Edinburgh, 2006.
[4] Wagner, P., “Great expectations–Introspective vs. perceptual prominence ratings and their acoustic correlates”, In Ninth European Conference on Speech Communication and Technology, ISCA, 2005.
[5] Mo, Y., Cole, J. and Hasegawa-Johnson, J., “How do ordinary listeners perceive prosodic prominence? Syntagmatic vs. paradigmatic comparison”, Poster presented at the 157th Meeting of the Acoustical Society of America, Portland, Oregon, 2009. Columbus, OH: Department of Psychology, Ohio State University, 2007. Retrieved March 15, 2006, from www.buckeyecorpus.osu.edu.
[6] Cole, J., Mo, Y. and Hasegawa-Johnson, M., “Signal-based and expectation-based factors in the perception of prosodic prominence”, Laboratory Phonology, 1, 425452, 2010.
[7] Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and et al., “Buckeye corpus of conversational speech (2nd release)”.
[8] Mo, Y., Cole, J. and Lee, E.K., “Naive listeners prominence and boundary perception”, Proc. Speech Prosody, Campinas, Brazil, 735–738, 2008.
[9] Arnold, J.E., Kam, C. L. H., Tanenhaus, M.K. and Arnold, J., “If you say thee uh-you’re describing something hard: The on-line attribution of disfluency during reference comprehension”, Journal of Experimental Psychology: Learning, Memory, and Cognition, vol 33, 5, 914-930, 2007.
[10] Moschou, V., Kotti, M., Benetos, E. and Kotropoulos, C., “Systematic comparison of BIC-based speaker segmentation systems”, IEEE Workshop on Multimodal and Multimedia Signal Processing, 2007.

This study is supported by NSF IIS-0703624 to Cole and Hasegawa-Johnson