

Dynamic Measures of Stance-taking in Conversational Speech

Gina-Anne Levow, Richard A. Wright

Department of Linguistics, University of Washington

levow, rawright@uw.edu

Stance, the expression of a speaker’s attitudes or opinions about the topic under discussion, plays an important role in negotiation and decision-making. While stance-taking behavior has received moderate attention in text, the correlates of attitude-expression in speech have been less well-studied.

We build on the ATAROS corpus (Freeman et al., 2014), a corpus of dyadic task-oriented conversations designed to elicit high rates of stance-taking at varying levels of strength and involvement. Focusing on the tasks designed to elicit lower and higher levels of involvement, a 20-dyad subset of this corpus has further been annotated for stance strength (low to high), polarity (positive to negative), and act type (agreement, opinion-offering, convincing, etc) (Freeman, 2015). Prior analysis has identified significant differences in several acoustic-prosodic measures across these dimensions of stance. As with much prior work, these contrasts have been expressed in terms of maxima, minima, ranges, means of pitch, intensity, and duration. However, measures that capture the dynamics of acoustic-prosodic phenomena in speech have been less studied.

In this work, we consider two such measures: modulation spectrum (Shinozaki et al, 2009) and pitch change (Slaney et al, 2013). The former analyzes spans of speech by treating the energy contour as a signal and characterizing its dynamics through Fourier analysis. Modulation spectrum as a measure of speaking style has been shown to effectively distinguish adult-directed and child-directed speech, improving over typical measures of pitch and intensity. The latter avoids error-prone pitch prediction in favor of tracking the dynamics of pitch change by exploiting autocorrelation across subbands. The pitch change measures have been successfully employed to improve automatic recognition of lexical tone in Mandarin Chinese, improving over standard pitch tracking.

The modulation spectrum captures modulation over longer spans of speech, so we apply it to investigate coarse-grained phenomena of stance strength, annotated over spurts (spans of speech delimited by at least 500ms of silence). We compute measures over two bands: 0-3kHz and 3-6kHz. We find significant differences in modulation spectrum measures for the different levels of stance strength in our corpus. In contrast, pitch change is computed at 10ms intervals, and thus is applicable to both coarse and fine-grained stance-taking behaviors. At the coarse-grained spurt-level, all levels of stance strength differ significantly under measures of pitch change. In particular, greater levels of steeper pitch increase are associated with greater stance strength. In the case of fine-grained stance acts, we consider only the nine most common individual acts with at least 80 occurrences in our corpus. We also find that stance type exhibits a significant effect on measures of pitch change. These findings indicate that these measures of speech dynamics can help to characterize speaking style related to stance expression in conversational collaborative tasks.

References

Freeman, V. (2015). The phonetics of stance-taking. Doctoral dissertation, University of Washington.

Freeman, V., Chan, J., Levow, G.-A., Wright, R.A., Ostendorf, M., and Zayats, V. (2014). Manipulating stance and involvement using collaborative tasks: An exploratory comparison. Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech 2014), Singapore, Sept. 14-18.

Shinozaki, T., Ostendorf, M., and Atlas, L. (2009). Characteristics of speaking style and implications for speech recognition. *Journal of Acoustical Society of America*, 126(3):1500–10.

Slaney, M., Shriberg, E. and Huang, J.T. (2013) Pitch-Gesture Modeling Using Subband Autocorrelation Change Detection. Proceedings of Interspeech 2013, ISCA, 29 August 2013.