Investigating an Acoustic Measure of Perceived Isochrony

Shannon Mooney (*Georgetown University*) and Grace C. Sullivan (*Georgetown University*)

In a preliminary investigation of isochrony, the rhythmic integration of talk, we evaluated rhythmic phenomena previously theorized to coordinate turn-transitions (Scollon 1982) for correlates in the acoustic signal. Rhythmic sequencing is one of many elaborate "contextualization cues" (Gumperz 1982) regarded as facilitating a successful turn-transition (Erickson 1982a,b). Previous studies of rhythm in conversation have attended only to its perceptual and interactional facets. In addressing this gap, our study finds quantitative justification for such claims of rhythmic turn-taking.

We selected for acoustic analysis in Praat (Boesma & Weenick 2014) the twelve non-task based, dyadic conversations of the *Santa Barbara Corpus of Spoken American English* (*SBCSAE)*. This process involved automatic phonemic alignment of the conversations using FAVE (Rosenfelder et al. 2014). Following Marcus's (1981) assertion that the onset of the vowel is the closest acoustically-measurable location to the perceptual center of the syllable where the rhythmic downbeat occurs, duration was measured between vowel onsets to create prosodic syllables. Not all prosodic syllables can contain a rhythmic beat, and those that can are characterized as "prominent" in nature (Couper-Kuhlen 1993). Perceptual "prominence" of a syllable can be difficult to determine based on acoustic information alone, as it can possibly correlate with measures such as relative duration, amplitude, pitch, stress, and weight. Of all these measures, Couper-Kuhlen reports that the strongest correlation exists between perceived rhythmic beat locations and long duration prosodic syllables (1993: 19), thus for this preliminary analysis we attributed prominence to prosodic syllables longer than each conversation's mean syllable duration. Out of 42,807 prosodic syllables measured, this yielded 15,972 prominent prosodic syllables.

The units of duration between prominent syllables, hereafter intervals, were judged to form an isochronic sequence when the durations between at least three consecutive intervals varied by less than the conservative measure of the perceptual threshold for tolerance of isochrony, up to a 30% variance (Couper-Kuhlen 1993: 78). This measure revealed 564 rhythmic sequences across the twelve *SBCSAE* conversations, which ranged in duration between one and ten seconds and consisted of up to eleven intervals. Of these, 208 or 37% appeared within turn-transitions, and results from our preliminary analysis indicated that rhythmic sequencing was significantly more likely to appear within a turn-transition than outside of one ($p<.01$). Our analysis shows that isochrony is not simply perceptual in nature, but that it has a quantifiable correlate in the acoustic signal.

Our findings of significant isochrony in the turn-transitions of the *SBCSAE*, a corpus often used in discourse analysis, confirms what interactional sociolinguists such as Scollon and Erickson have long argued-- that rhythmic cues aid the coordination of talk between speakers in turn-transitions. We can confirm that these rhythmic cues are a component of turn-transitions not only perceptually, but acoustically as well. Future directions of this project are investigation of other acoustic measures of prominence and speaker perception of acoustic rhythm.

Abercrombie, D. 1968: Some functions of silent stress. Work in Progress 2, Edinburgh
University Department of Linguistics.

Boersma, P., & Weenink, D. 2014. Praat: doing phonetics by computer [Computer program].
Version 5.3.84.

Couper-Kuhlen, E. 1993. English speech rhythm: Form and function in everyday verbal
interaction (Vol. 25). John Benjamins Publishing.

Du Bois, W., Chafe, W.L., Meyer, C., Thompson, S.A., Englebretson, R., and Martey N. 2000-
2005. Santa Barbara Corpus of Spoken American English, Parts 1-4. Philadelphia:
Linguistic Data Consortium.

Erickson, F. 1982a. Money tree, lasagna bush, salt and pepper: Social construction of topical
cohesion in a conversation among Italian Americans. In Tannen, ed. Analyzing
Discourse: Text and Talk. Washington, D.C.: Georgetown Univ. Press, 43-70.

Erickson, F., & Shultz, J. 1982b. The counselor as gatekeeper: Social interaction in interviews.
New York: Academic Press.

Goffman, E. 1981. Forms of Talk. University of Pennsylvania Press.

Gumperz, J. J. 1982. Discourse Strategies. Cambridge: Cambridge University Press.

Kendall, T. S. 2009. Speech rate, pause, and linguistic variation: An examination through the
sociolinguistic archive and analysis project. Dissertation.

Marcus, S. M. 1981. Acoustic determinants of perceptual center (P-center) location. Perception
& Psychophysics, 30(3), 247-256.

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H. and Yuan, J.
2014. FAVE 1.1.3. ZENODO. doi:10.5281/zenodo.9846

Sacks, H., Schegloff, E. A., and Jefferson, G. 1974. A simplest systematics for the organization
of turn-taking for conversation. Language, 696-735.

Scollon, R. 1982. The rhythmic integration of ordinary talk. In Tannen, ed. Analyzing Discourse:
Text and Talk. Washington, D.C.: Georgetown Univ. Press, 335-349.

Wise, A. 2008. sbc_to_tg.pl. http://ncslaap.lib.ncsu.edu/tools/sbc_to_praat.php