Prosodic Convergence During and After a Cooperative Maze Task

Yoonjeong Lee[1]*, Samantha Gordon Danner*[1]*, Benjamin Parrell*[2]*, Sungbok Lee*[1]*, Louis Goldstein*[1]*, Dani Byrd*[1]
[1]University of Southern California, [2]University of Delaware
yoonjeol@usc.edu, sfgordon@usc.edu

Convergence effects have been found in many aspects of communication between speakers in conversation. Previous research has used conversations elicited during cooperative gameplay or simultaneous speech tasks to elicit broad measures of convergence such as speech rate, pitch, intensity and kinematic trajectories (e.g., Levitan & Hirschberg, 2011; Tiede et al., 2010). More recent work (Gravano et al., 2014) has also found convergence effects in *qualitative* measures, like occurrence and frequency of prosodic contours, but prior work has not addressed the relationship between convergence and *qualitative* and *quantitative* measures of prosody. Moreover, little is known about whether convergence effects are maintained after conversation has concluded. Here, we are interested in convergence in both qualitative and quantitative aspects of prosodic structure, and we measure whether these effects persist in individual speech tasks after conversation has ended. Our work tests the following research questions: 1) How do prosodic properties such as boundary strength, choice of boundary tone and the quantitative properties of those tones differ when a speaker is engaged in an individual speech task versus when two speakers work on a cooperative speech task? 2) If speakers converge on prosodic patterns during the cooperative task, is the pattern maintained after the task has concluded?

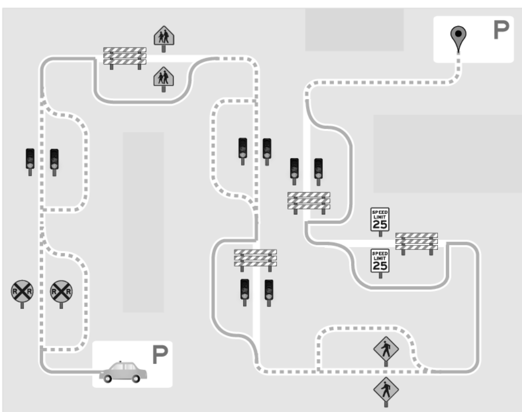

Figure 1: Cooperative maze example image

In the present study, two WAVE electromagnetic articulography systems (Northern Digital, Inc.) were used simultaneously to record speech articulator movements (not reported here at this time) from a dyad. Acoustic data was collected for two female native speakers of American English, F1 and F2, not previously known to each other. Speakers were seated at desks approximately 2 meters apart, facing each other with a clear line of sight. Each speaker had a computer monitor on their desk to view instructions. During non-cooperative portions of the experiment, one speaker completed an individual speech task while the other speaker listened to music through over-ear headphones, and a large board was placed between the speakers to prevent their view of each other. Each speaker individually completed a maze task, following a highlighted path through eight landmarks per maze. Participants described each landmark, using sentences of the form '*And then you go* [between/beside] *the next two* [lights/signs].' The barrier and headphones were removed for the cooperative maze task. Each speaker saw a different version of a set of mazes that only included the portion of the maze to be navigated by that speaker (Fig. 1). Each speaker navigated either one or two consecutive landmarks on a given turn before turning over the floor to the other speaker for the next maze section. Following the cooperative maze task, each speaker repeated the individual maze task.

We examined temporal and intonational properties of each speaker's prosodic structure before, during and after the cooperative maze task. We measured 1) sentence durations (to replicate findings showing convergence of speech rate); 2) *FL* – final lengthening, as described below in (a), as a measure of prosodic boundary strength; and 3) occurrence frequency of sentence-final boundary tones (L% or H%), and the value of low and high F0 peaks in sentence-final target words ('*lights*' or '*signs*') (Beckman & Elam, 1997).

$$\text{(a) Final lengthening (FL)} = \frac{\text{sentence-final target word duration (ms)}}{\text{sentence duration (ms)}}$$

Our speakers exhibited distinct temporal and intonational patterns *before* the cooperative maze task. Average sentence durations were much shorter for speaker F1 than F2 (sentence: mean diff. 669 ms). *FL* values were higher in F2's speech compared to F1's (F1_FL: .27 vs. F2_FL: .24). Additionally, while F1 always showed a rising pattern at the end of the sentence (H%), F2 preferred a falling intonation or no rising (F2_L%: 62%). The mean F0 for F1 was higher and had a wider range than F2's mean F0.

Our results show that many phonetic properties of F2's speech became more similar to F1's speech properties over the course of the experiment; however in contrast, F1 did not show any significant prosodic modification during or after the cooperative task. F2's sentence duration became shorter and more comparable to F1's (mean diff. 127 ms) during the cooperative task. This speech rate convergence effect for F2 carried on, to a lesser extent, *after* the cooperative task (mean diff. 247 ms). Similar patterns were found for F2's *FL* values (during & after: F1_FL: .26 vs. F2_FL: .28), indicating that F2 showed convergence to F1 in the dimension of prosodic boundary strength as well.

While F1 consistently used a rising boundary tone (F1_H%: 99.3%) in the cooperative maze task, F2 shifted her boundary tone type during (F2_H%: 83.1%), and to a lesser extent, after (F2_H%: 55.6%) the cooperative task (Fig. 2). These results are likely driven by the target word *signs*, which F2 continued to produce with an H% boundary even after the cooperative maze task: F2's H% occurrence for *signs* was significantly sparser in the first round of the individual maze task (68.8%) than during the cooperative maze task (98.6%) or during the subsequent individual maze task (100%).

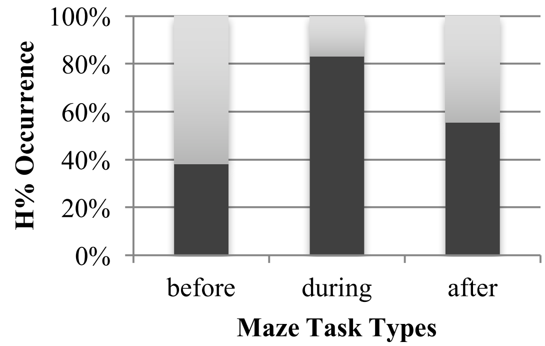

Figure 2: Boundary tone type shift in F2

F1 did not show any significant difference in peak F0 values of sentence-final target words produced with a rising boundary tone, across different maze tasks (average F1's H target value: 216 Hz). Conversely, F2 produced higher H targets during the cooperative maze task than before the cooperative task (mean diff. 57 Hz). This convergence effect persisted *after* the cooperative task (mean diff. 35 Hz before vs. after the cooperative task; see Fig. 3). F1 produced too few L% tokens to compare speakers' L targets.

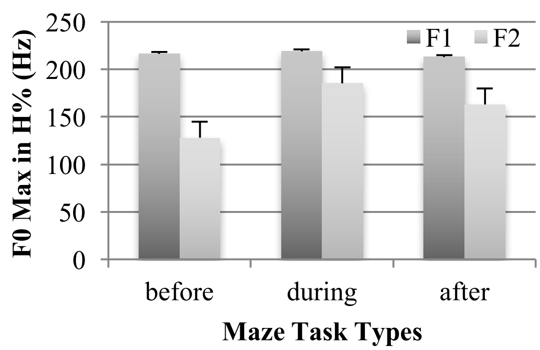

Figure 3: Convergence effect in H targets

In this work, we extend previous research on phonetic convergence in conversation. We found that convergence effects can persist even after a conversational speech interaction between speakers has ended. Moreover, we show that there were convergence effects in prosodic boundary strength, use of boundary tones, and quantitative properties of those boundary tones. Our study suggests that conversing speakers converge not only in broad measures like speech rate, but also in qualitative and quantitative aspects of prosodic structure. These findings raise the question of whether quantitative aspects of prosody – which can be produced by adaptation of speech motor behavior – and qualitative aspects of prosody – which engage the grammar and social interaction – can be modeled as arising from a single source.

References

Beckman, M. E., & Elam, G. A. (1997). *Guidelines for ToBI Labelling*. || Gravano, A., Benus, S., Levitan, R., & Hirschberg, J. (2014). Three ToBI-Based Measures of Prosodic Entrainment and their Correlations with Speaker Engagement. In *Spoken Language Techonlogy Workshop*, 578–583. || Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *INTERSPEECH*, 3081–3084. || Tiede, M., Bundgaard-Nielsen, R., Kroos, C., Gibert, G., Attina, V., Kasisopa, B.,Vatikiotis-Bateson, E., & Best, C. (2010). Speech articulator movements recorded from facing talkers using two electromagnetic articulometer systems simultaneously. *Proceedings of Meetings on Acoustics*, *11*, 060007–060009.