Prosodic Entrainment and Tutoring Dialogue Success

Jesse Thomason, Huy Nguyen, & Diane Litman

University of Pittsburgh Computer Science Department
Pittsburgh, PA 15213

Abstract

This study investigates the relationships between student entrainment to a tutoring dialogue system and learning. By finding the features of prosodic entrainment which correlate with learning, we hope to inform educational dialogue systems aiming to leverage entrainment. We propose a novel method to measure prosodic entrainment and find specific features which correlate with user learning. We also find differences in user entrainment with respect to tutor voice and user gender.

Motivation and Goals

- Entrainment occurs when speakers unconsciously mimic one another's voices, diction, and other behaviors [2]
- In tutoring dialogues, entrainment has been found to correlate with learning gain [7] and reflect users' emotional states [4]
- Spoken dialogue systems might be able to encourage learning by entraining or encouraging entrainment
- To what features of a student's voice should a system accommodate if using this strategy?
- We examine 29 students' dialogue with the ITSPOKE [1] tutoring dialogue system; the tutor voice was either a pre-recorded or synthesized voice for each student
- We considered the mean, min, max, and standard deviation of the speech signal amplitude (RMS) and pitch (F0) of every utterance
- We find features of students' voices correlated with learning and demonstrate a new metric for measuring entrainment

Hypotheses

We hypothesized we would find that entrainment

1. *positively correlated with learning gain*. Past literature suggests correlations with both learning [4, 7] and task success [3].
2. *was higher for students interacting with the pre-recorded tutor voice*. This would inform a system that elicits entrainment or accommodates.
3. *was higher for males*. Psychological research suggests that males entrain more than females when they are in a subservient role of conversation [5]. A system utilizing entrainment may need to consider student gender.

Entrainment Metrics

Avg

Proposed and used by [3], this metric defines entrainment between the student s and tutor t on speaker-normalized feature f, where speaker$_f$ is the speaker's mean for f over the dialogue, as

$$ent(s,t) = -|s_f - t_f|$$

Reg

Each tutor-student exchange was plotted as a point. The similarity r^2 of the linear regression between tutor and student feature values was calculated for the 3rd through final exchange of the dialogue. The entrainment score was calculated as the correlation coefficient r of the regression between these similarity scores and the number of exchanges that took place to form them, as demonstrated in the figure below for the RMS Max feature.

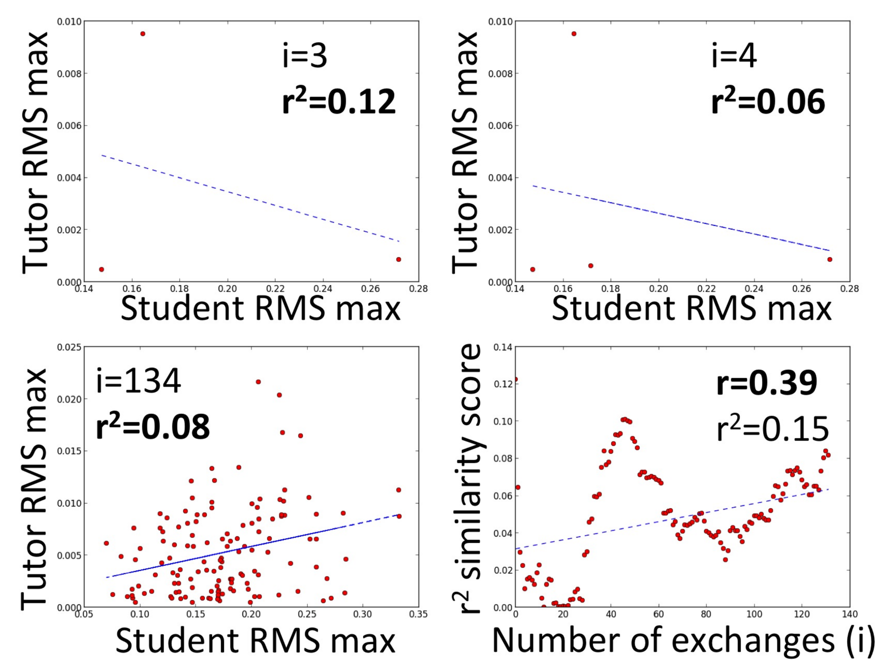

Figure: Visualizing the **Reg** entrainment metric

Results

We found all significant correlations between our calculated entrainment scores and normalized learning gain, $NLG = \frac{post-pre}{1-pre}$. As in [7], we examined students in high- and low-pretest groups as well, splitting by the median score.

Table: Hypothesis 1. * denotes significance ($p < 0.05$), while $^+$ denotes a trend ($p < 0.1$).

Group	Metric	Direction	Entrainment
all	Avg	↗	F0 Min$^+$, F0 Max$^+$, F0 Stddev$^+$
low	Reg	↗	RMS Min*
high	Avg	↗	F0 Mean*, F0 Stddev*
high	Reg	↗	F0 Max*

We used Welch's two-tailed t-tests to determine if there were significant differences between users' mean entrainment in the pre-recorded (15 students) and synthesized (14 students) voice conditions or between male (12 students) and female (17 students) mean entrainment.

Table: Hypotheses 2 and 3. * denotes significance ($p < 0.05$), while $^+$ denotes a trend ($p < 0.1$).

Metric	Direction	Entrainment
Avg	pre$>$syn	F0 Stddev$^+$
Reg	pre$>$syn	F0 Mean$^+$, F0 Min$^+$
Avg	male$>$female	RMS Max*, RMS Min*

Discussion and Future Work

Returning to our hypotheses, our results suggest the following.

1. *support*. Learning gain positively correlated with entrainment for several pitch features when considering all students, significantly so for high-pretesters alone, and for the loudness min feature significantly so for low-pretesters alone.
2. *partial support*. The means of several pitch entrainments in the pre-recorded condition were found higher than those in the synthesized condition.
3. *support*. Male mean entrainment was significantly higher than female mean entrainment on loudness min and max features.

When comparing our proposed **Reg** metric to the existing **Avg** metric, we note that:

- **Reg** does not require normalization and can be deployed in a live system
- **Reg**, which captures similarity over time, and **Avg**, which measures average dialogue similarity, detect complementary sets of entrained features

The next steps in this work will be to:

- Analyze differences between our new entrainment metric and other established metrics
- Explore lexical entrainment
- Investigate finer-grained entrainment calculations

References

[1] K. Forbes-Riley, D. Litman, S. Silliman, and J. Tetreault.: Comparing Synthesized versus Pre-Recorded Tutor Speech in an Intelligent Tutoring Spoken Dialogue System. In: Proc. 19th International Florida Artificial Intelligence Research Society, pp 509-514. Melbourne Beach, FL (2006)

[2] J. L. Lakin, V. E. Jefferies, C. M. Cheng, and T. L. Chartrand.: The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. Springer Journal of Nonverbal Behavior 27(3) 145-162 (2003)

[3] R. Levitan, A. Gravano, L. Willson, S. Benus, J. Hirschberg, A. Nenkova.: Acoustic-Prosodic Entrainment and Social Behavior. In: Proc. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT), pp11-19. ACM, Montreal, Canada (2012)

[4] C. M. Mitchell, K. E. Boyer, and J. C. Lester.: From Strangers to Partners: Examining Convergence within a Longitudinal Study of Task-Oriented Dialogue. In: Proc. 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp94-98. ACM, Seoul, South Korea (2012)

[5] J. S. Pardo.: On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America 119(4) 2382-2393 (2006)

[6] A. Raux and M. Eskenazi.: Non-Native Users in the Let's Go!! Spoken Dialogue System: Dealing with Linguistic Mismatch, pp217-224. In: Proc. NAACL HLT (2004)

[7] A. Ward and D. Litman.: Dialog convergence and learning. In: Proc. 13th International Conference on Artificial Intelligence Education. Los Angeles, CA (2007)