Prosodic Entrainment and Tutoring Dialogue Success

Jesse Thomason, Huy V. Nguyen, and Diane Litman

University of Pittsburgh, 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.

1 Introduction

Spoken dialogue systems offer students one-on-one instruction from a computer tutor. Entrainment occurs when speakers unconsciously mimic one another's voices, diction, and other behaviors [2]. In tutoring dialogues, [7] found entrainment from students with high pre-test scores correlated with learning gain, and [4] found such correlations to learning and negative emotional states. If a system encouraged entrainment from users, as the system in [6] did to improve speech recognition, it might reduce negative states and encourage learning.

Knowing which entrainment features are correlated with learning gain would inform this strategy. We searched an existing intelligent tutoring dialogue system corpus to find such correlations with speech features. There is no standard for measuring prosodic entrainment, though several methods exist. We calculated entrainment with both a recent metric [3] and a new metric we propose.

2 Data and Post-hoc Experiment

Our data comes from an experiment using the ITSPOKE tutoring dialogue system [1]. Each student interacted with either a pre-recorded or synthesized tutor voice. They verbally responded to tutor questions for 5 problem dialogues over one or more sessions. Pre- and post- test scores were recorded. We considered only students who experienced no technical problems, and completed all problem dialogues and a post-experiment survey, which gave us 29 total students. 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.* If true, this would inform a system that elicits entrainment or accommodates.

K. Yacef et al. (Eds.): AIED 2013, LNAI 7926, pp. 750–753, 2013.

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.

2.1 Entrainment Features

We used openSMILE[1] to extract prosodic features. Specifically, we considered the mean, min, max, and standard deviation of the speech signal amplitude (RMS) and pitch (F0) of every utterance. We captured entrainment on each feature f in two ways. In each, we consider the pre-recorded and synthesized tutor voices as their own speakers.

In the first method, we speaker-normalized each feature value via z-scores and used the metric proposed by [3]. In our domain, it defines entrainment between the student s and tutor t on feature f as $ent(s,t) = -|s_f - t_f|$ where speaker_f is the speaker's mean for f over the dialogue. We denote this entrainment calculation metric **Avg**.

Additionally, we proposed a metric to capture changes in exchange-level similarity throughout a dialogue. For each student s, we divided the dialogue into N consecutive exchanges. Each exchange was a pair of student/tutor utterances where the student s was directly responding to the tutor t. These formed a sequence of exchanges $(n_1, \ldots, n_N)$ where each $n_i = (f_{ti}, f_{si})$, the tutor and student raw feature values on the turns of exchange i. We denote the sequence of the tutor's feature values from the i to jth exchange as $T_i^j = (f_{ti}, f_{ti+1}, \ldots, f_{tj})$ and the student's as $S_i^j = (f_{si}, f_{si+1}, \ldots, f_{sj})$. We give a similarity score which considers preceding exchanges[2] when scoring the current exchange. Specifically, we define $\text{sim}(j) = \text{linreg}_{r^2}(T_3^k, S_3^k), 3 \leq k \leq j$, where linreg_{r^2} is the fit coefficient r^2 of a linear regression between the two sequences. We calculate the entrainment on f for this student/tutor pair as $ent(s,t) = \text{linreg}_r(j, \text{sim}(j)), 3 \leq j \leq N$, where linreg_r is the fit coefficient r of the linear regression between the similarity scores and the number of consecutive exchanges that yielded them. Figure 1 outlines this calculation. We expect $ent(s,t)$ to be more positive on feature f when the student is converging to the tutor's feature f values over the course of the dialogue. We denote this entrainment calculation metric **Reg**.

2.2 Experimental Methods and Results

We judged student learning using normalized learning gain, $NLG = \frac{post-pre}{1-pre}$, then found all significant correlations between our calculated entrainment scores and learning. As in [7], we performed correlation tests for students in high- and low-pretest groups as well. We divided these groups by the median pre-test score (students with median score were not considered). Table 1 summarizes the correlations found between entrainment[3] and learning in these groups.

[1] `http://opensmile.sourceforge.net/`

[2] We start with 3 exchanges because a regression is trivial on 2 and undefined on 1.

[3] We denote entrainment scores for a feature by that feature's abbreviated name. Thus, *RMS Max* denotes the entrainment values for the loudness maximum feature.

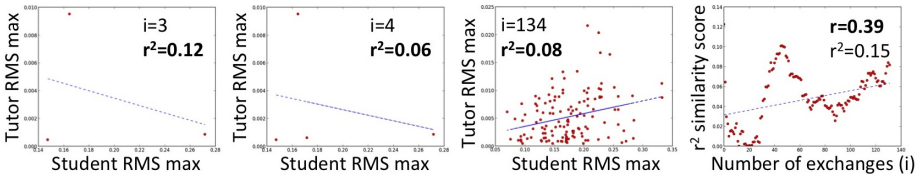

Fig. 1. Each tutor-student exchange was plotted as a point. The similarity r^2 of the linear regression between tutor and student was calculated for the 3rd through Nth exchange. The entrainment score was calculated as the correlation coefficient r of the regression between these similarity scores and the number of exchanges that took place to form them.

Table 1. Correlations of student learning (NLG) with entrainment scores for all students and for low- and high-pretest groups. * 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 2 summarizes differences found between mean entrainments in those pairs.

Table 2. Differences in entrainment means between students in the pre-recorded versus synthesized condition and between male and female students. * 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*

3 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.

We support existing claims that entrainment correlates with student performance in intelligent spoken tutor dialogue systems. Our results suggest student entrainment correlates with learning and that tutor voice and gender both affect entrainment. Our new metric for capturing prosodic entrainment in a turn-taking scenario does not require normalization and could be deployed in a live system, unlike that of a recent work [3]. We find that the entrainment correlations it detects complement those detected by the metric used in [3]. Thus the new metric, which captures changes in similarity over time, might be useful in tandem with metrics similar to that of [3], which measure average dialogue similarity.

In the future, we will further analyze differences between our new entrainment metric and those established. We will also explore lexical entrainment. Students may reset their entrained behaviors on new problems or new sessions with the tutor, so we will investigate finer-grained entrainment calculations.

Acknowledgments. We thank the ITSPOKE group for their helpful feedback and the reviewers for their suggestions.

References

1. Forbes-Riley, K., Litman, D., Silliman, S., Tetreault, J.: Comparing Synthesized versus Pre-Recorded Tutor Speech in an Intelligent Tutoring Spoken Dialogue System. In: Proc. 19th International Florida Artificial Intelligence Research Society, Melbourne Beach, FL, pp. 509–514 (2006)
2. Lakin, J.L., Jefferies, V.E., Cheng, C.M., Chartrand, T.L.: 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. Levitan, R., Gravano, A., Willson, L., Benus, S., Hirschberg, J., Nenkova, A.: 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), pp. 11–19. ACM, Montreal (2012)
4. Mitchell, C.M., Boyer, K.E., Lester, J.C.: 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, pp. 94–98. ACM, Seoul (2012)
5. Pardo, J.S.: On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America 119(4), 2382–2393 (2006)
6. Raux, A., Eskenazi, M.: Non-Native Users in the Let's Go!! Spoken Dialogue System: Dealing with Linguistic Mismatch. In: Proc. NAACL HLT, pp. 217–224 (2004)
7. Ward, A., Litman, D.: Dialog convergence and learning. In: Proc. 13th International Conference on Artificial Intelligence Education, Los Angeles, CA (2007)