# Building Rapport through Dynamic Models of Acoustic-Prosodic Entrainment

**Nichola Lubold**
Arizona State University
Tempe, AZ 85281 USA
Nichola.Lubold@asu.edu

## Abstract
As dialogue systems become more prevalent in the form of personalized assistants, there is an increasingly important role for systems which can socially engage the user by influencing social factors like rapport. For example, learning companions enhance learning through socio-motivational support and are more successful when users feel rapport. In this work, I explore social engagement in dialogue systems in terms of acoustic-prosodic entrainment; entrainment is a phenomenon where over the course of a conversation, speakers adapt their acoustic-prosodic features, becoming more similar in their pitch, intensity, or speaking rate. Correlated with rapport and task success, entrainment plays a significant role in how individuals connect; a system which can entrain has potential to improve social engagement by enhancing these factors. As a result of this work, I introduce a dialogue system which can entrain and investigate its effects on social factors like rapport.


CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3027132

## Author Keywords
dialogue; adaptive; acoustic-prosodic; entrainment

## ACM Classification Keywords
H.5.2 – User Interfaces (Voice I/O, Natural Language)
H.1.2 – User/Machine Systems

## Introduction
Dialogue systems have entered mainstream society in the form of personalized assistants like Siri and Cortana. We can imagine a future where we talk every day to our computers, where virtual agents provide therapeutic and emotional support, or act as virtual peers in the classroom, facilitating learning. Advances in automated speech recognition and dialogue management have made this future, where systems

can hold a conversation on par with a human dialogue partner and form social relationships, a realistic possibility. However it is an open question whether and how dialogue systems should form social relationships. In human-human interactions, successfully building and maintaining rapport plays a critical role in helping collaborating partners achieve task success [1]. Early evidence suggests similar processes are at work in human-computer interactions, where humans who feel rapport are more motivated, feel more self-efficacy and collaborate better [2, 3]. Dialogue systems which can build and maintain rapport have the potential to personalize and enhance interactions.

This research explores how we can enable dialogue systems to build and maintain rapport, particularly through acoustic-prosodic entrainment. Acoustic-prosodic entrainment occurs when conversational partners adapt their acoustic-prosodic speech features, such as pitch or intensity, over the course of a conversation. Correlated with rapport, communicative success, and likability [4, 5, 6], acoustic-prosodic entrainment can play a significant role in how conversational partners connect [7]. Replicating phenomena like entrainment in dialogue systems has potential for influencing rapport.

To explore the effects of acoustic-prosodic entrainment in dialogue, I frame this research within the context of peer tutoring because of the potential benefits building rapport holds within this domain. In peer-tutoring, tutors build knowledge by identifying tutee misconceptions and constructing explanations. In human-human peer-tutoring, rapport is correlated with learning, potentially because rapport leads students to encourage and challenge each other [1,8].

Transitioning the concepts of peer-tutoring to educational technology, learning companions can motivate and engage students by applying the same principles. For this work, the learning companion is a robot students teach how to solve math-based problems. Within the context of this activity, increased human-agent rapport might improve learning.

The goal of this research is to investigate how acoustic-prosodic entrainment can be used to increase the social engagement of dialogue systems by influencing social factors like rapport. As a result of this work, I introduce a dialogue system for a learning companion which has the ability to adapt its prosody in response to a user's, exploring effects on social factors.

## Background

Explored in-depth in human-human conversation, entrainment is both continuous (occurring consistently) and dynamic (users will entrain and then reset) as well as global (conversation wide) and local (turn-by-turn). Understanding human-human entrainment can inform the design of automated entrainment. Turn-by-turn, dynamic entrainment appears to be more consistently related to social factors; in my own prior work on turn-level entrainment, individuals entraining on pitch on a turn-by-turn basis have higher measures of communicative success [4] and rapport [5].

In relation to human-computer interaction, prior work has explored effects of static voices which match individual traits such as personality; for example, Nass and Brave found introverts prefer introverted voices [9]. Only recently has dynamic prosodic adaptation been explored; Levitan and colleagues [10] found people unconsciously trusted a virtual avatar which

adapted to the user's speaking rate and intensity more than one that did not. These results suggest dialogue systems which have the ability to entrain can influence social factors; this work builds on these insights and seeks to establish generalizable methods of adaptation given the complexity of human-human entrainment.

## Research Questions and Methods

Because individuals can entrain on many different kinds of acoustic-prosodic features (i.e. pitch, intensity, speaking rate) in many ways (growing closer, matching on a turn-by-turn basis), it is unclear how a system should entrain and the individual effects an entraining system might have. I pose the following research questions to ground my exploration and design:

> **RQ1**: What are the optimal methods for modeling human-human entrainment automatically?
>
> **RQ2**: How do different forms of automated entrainment affect social responses like rapport?

Answering these questions begins with an in-depth literature review on acoustic-prosodic entrainment in human-human interactions and data collection from human-human experimental sessions. Analyzing this data, we then develop potential models of entrainment, and apply these models to the dialogue system of a learning companion. Experimental evaluation differentiates which methods or models of human-human entrainment are the most effective and subsequent effects on social responses.

## Results & Next Steps

Utilizing prior findings in human-human entrainment and my own analysis of human-human corpora, I have identified two potential models for entrainment which I am in the process of evaluating. The first model targets individual features for adaptation; this model is based on the theory that dynamic, turn-by-turn entrainment is more indicative of social factors and its effects can be broken down by individual features. The second model is a holistic view of entrainment; based on findings that entrainment typically occurs on multiple features simultaneously, this model focuses on entrainment across features and finding recurring patterns.

I have explored the first model utilizing pitch. In my prior work, entrainment on pitch had the highest relationship to rapport and communicative success [4, 5]. I identified and evaluated different types of pitch adaptations and found that adapting to a user's mean pitch could achieve higher 3rd party perceptual ratings of naturalness and rapport over other types of pitch adaptions. I also found that while the pitch adaptation resulted in more rapport and naturalness, it was not significantly better than normal text-to-speech output. In further experimental evaluations focusing on how individuals express rapport when engaging with a pitch-adaptive interface, males, when compared to females, responded significantly more positively to the adaptation, evincing increased rapport-building responses.

The first model resulted in mildly significant effects, but my findings suggest that individual differences in perception and degree of entrainment play a very pertinent role. I am currently in the process of developing the second model and am collecting additional human-human data for additional pattern recognition and training. Utilizing probabilistic models including neural networks, I am identifying patterns of entrainment in high-rapport dyads; these patterns will

be applied to the dialogue system of the learning companion and evaluated.

## Contributions

While any discovered effects of the proposed dialogue system will be limited to the context of peer tutoring, the adaptive dialogue system itself can transition to other domains; the process of analyzing entrainment patterns and training a dialogue system based on these patterns will be useful for the future development of socially engaging dialogue systems. In addition, providing an understanding of how individual differences such as gender might play a role in social responses to adaptive dialogue systems can act as a guide for future innovation. Finally, exploring the effects of replicating phenomena like entrainment may provide insight into how to increase social factors like rapport and in the case of peer-tutoring, fundamental factors like learning in educational technologies.

## Acknowledgements

This work is advised by Erin Walker, Assistant Professor in Computer Science at Arizona State University, and Heather Pon-Barry, Assistant Professor in Computer Science at Mount Holyoke College. This work is supported in part by the Ira A. Fulton Schools of Engineering through a Dean's Fellowship and the Google Anita Borg Memorial Scholarship.

## References

1. Ogan, Amy, et al. "Rudeness and rapport: Insults and learning gains in peer tutoring." *International Conference on Intelligent Tutoring Systems*. Springer Berlin Heidelberg, 2012.
2. Szafir, Daniel, and Bilge Mutlu. "Pay attention!: Designing adaptive agents that monitor and improve user engagement." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012.
3. Kang, Sin-Hwa, Jonathan Gratch, and James H. Watt. "The effect of affective iconic realism on anonymous interactants' self-disclosure." *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2009.
4. Borrie, Stephanie A., Nichola Lubold, and Heather Pon-Barry. "Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges." *Frontiers in Psychology* 6 (2015).
5. Lubold, Nichola, and Heather Pon-Barry. "Acoustic-prosodic entrainment and rapport in collaborative learning dialogues." *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*. ACM, 2014.
6. De Looze, Céline, et al. "Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction." *Speech Communication* 58 (2014): 11-34.
7. Herbert H. Clark. *Using Language*. Cambridge: Cambridge University Press, 1996.
8. Robinson, Debbie R., Janet Ward Schofield, and Katrina L. Steers-Wentzell. "Peer and cross-age tutoring in math: Outcomes and their design implications." *Educational Psychology Review* 17.4 (2005): 327-362.
9. Nass, C. I., & Brave, S. *Wired for Speech: How voice activates and advances the human-computer relationship*. Cambridge: MIT press, 2005.
10. Levitan, R. "Entrainment in Spoken Dialogue Systems: Adopting, Predicting and Influencing User Behavior." *HLT-NAACL.* 2013.