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.

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 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 adaptations. 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 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. I am utilizing a neural network to identify 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.

Research Situation & Status
As a PhD candidate at Arizona State University, my research interests are in the intersection of human-computer interaction and spoken dialogue with applications to education. This is supported by my thesis co-chairs, Dr. Erin Walker (Arizona State University) and Dr. Heather Pon-Barry (Mount Holyoke College), who specialize in human-computer interaction and spoken language processing respectively. I started my PhD in August 2013 and anticipate completing my thesis in the summer of 2018. I will be completing all course-work requirements the fall of 2016 and defending my proposal this coming spring.

References
Appendix 1

From the Doctoral Consortium I hope to gain feedback on the general direction of my dissertation as I write and prepare to defend my proposal this coming spring. I have been pursuing a particular model in the development and design of an adaptive dialogue system and as I transition to another potential model, I would be very interested in the thoughts from the community on direction, what is the most interesting, and possible pitfalls. My research also crosses several areas and I think I could greatly benefit from feedback from such an interdisciplinary community.

In terms of how I can benefit the consortium, I hope that one of my contributions may be to spark interest and discourse around the future potential of dialogue systems as an interaction modality, because while dialogue systems have been growing more popular, traditionally they have an unfortunate reputation in the CHI community, as discussed by Munteanu and colleagues [1]. In addition, I have my own lessons learned from performing research like mine, where in my search for dialogue systems which can adapt to their users we’ve had some surprising discoveries. I would love the opportunity to share these and hope that others can learn from my own experiences. I also hope to contribute by providing feedback to my fellow students; with a background in computer engineering, a unique focus on speech, 5 years of work experience, and 3 years in my program, I believe I can draw on my experiences to provide rich feedback that is not just helpful but actionable.


Appendix 2

(Please see following page)
To whom it may concern:

As Nichola Lubold’s primary dissertation adviser and thesis committee co-chair, I can assert that she would be both a good fit for this doctoral consortium and the doctoral consortium would be a good fit for her. I am an Assistant Professor in the School of Computing, Informatics, and Decision Systems Engineering at Arizona State University. I have known Nichola since she began her graduate studies at ASU in Fall 2013, and have been advising her since Fall 2014.

Now in her fourth year, Nichola is currently working on her proposal, with the intent to defend it this coming spring. Her research centers around how to incorporate social content and prosody in spoken dialogue systems, and what effects these social systems have. Nichola has completed a large portion of the preliminary work for her proposal, and has established a record of publication in this area, with papers in HRI and ASRU last year. Attending the consortium would be an excellent opportunity for her to get new perspectives on her goals and research questions as well as get feedback on her overall direction before she finalizes her proposal. Nichola would also benefit from the opportunity to present and discuss this work with the CHI community, as she has had a significant focus on human-computer interaction in her work but has never attended CHI. The community of scholars she would meet at the consortium would be a great support network for her as she moves forward in her research.

Nichola would be able to contribute to the DC in a number of ways. Her expertise relates to fostering social engagement in a number of contexts (e.g., digital assistants, human-robot interactions, and interactive learning environments) and using a number of methods (ranging from machine learning to quantitative coding of human-human interactions. Thus, she is able to provide informed feedback in a diverse set of HCI subareas. In addition, she is collaborative and enjoys exchanges of ideas, and thus will likely be an active and enthusiastic participant in the discussion.

I strongly support Nichola’s application to the CHI DC. Please do not hesitate to contact me with further questions.

Sincerely,

Erin Walker
Assistant Professor
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
Appendix 3 (Curriculum Vitae)

Education

**Arizona State University**, Tempe, AZ
Ph.D., Computer Science, 2013 – 2018 (Expected), GPA: 3.9

**University of Notre Dame**, Notre Dame, IN
B.S., Computer Engineering, 2005 – 2008 (graduated in 3 years)

Publications


* Awarded Best Poster Presentation at SLT 2014


Fellowships & Honors
NCWIT Collegiate Award, Honorable Mention 2016
Google Anita Borg Memorial Scholarship 2015
Grace Hopper Celebration Scholar 2015
National Science Foundation Graduate Research Fellowship, Honorable Mention 2015
Dean’s Fellowship, Ira A. Fulton Schools of Engineering 2013
4 year fellowship, awarded to 4% of all ASU doctoral students
Six Sigma GreenBelt, General Electric 2010
Upsilon Pi Epsilon Honor Society, University of Notre Dame 2008
Engineering Honors Program, University of Notre Dame 2008
Engineering Scholars Program, University of Notre Dame 2005

Work Experience
Data Science Intern – Intuit  Summer 2016
Program Manager – GE Capital, Retail Bank 2011 - 2013
Technology Project Manager – GE Capital, Retail Bank 2010 - 2011
Information Technology Leadership Program – GE Capital 2008 - 2010

Research Experience
Research Assistant, ARIZONA STATE UNIVERSITY 2014 – PRESENT
HCI and educational systems, spoken language processing. Advisor: Erin Walker, CIDSE

Research Assistant, ARIZONA STATE UNIVERSITY 2013 - 2014
Dialogue systems, natural language processing. Advisor: Heather Pon-Barry, CIDSE

Research Assistant, ARIZONA STATE UNIVERSITY 2014
Educational systems, intelligent tutoring. Advisor: Kurt VanLehn, CIDSE

Research Assistant, UNIVERSITY OF NOTRE DAME 2006 - 2008
Biometrics & pattern recognition. Advisor: Kevin W. Bowyer, CS & Engineering

TEACHING EXPERIENCE
ASU 101, Introduction to ASU, Instructor (135 students) – ARIZONA STATE UNIVERSITY 2016

FSE 104/494, EPICS, Faculty Associate (30 students) – ARIZONA STATE UNIVERSITY 2015 – Present

CSE 110, Introduction to Java, Instructor (117 students) – ARIZONA STATE UNIVERSITY 2014

CSE 476 / 598, Introduction to Natural Language Processing, TA – ARIZONA STATE UNIVERSITY 2013