Two Partners, One Voice: Prosodic Speech Entrainment in Romantic Relationships

by
Sarah Weidman

A Thesis

Presented to the Faculty of Mount Holyoke College

in partial fulfillment of the requirements

for the degree of Bachelor of Arts

with Honors

Department of Psychology and Education

Mount Holyoke College

South Hadley, Massachusetts

ACKNOWLEDGEMENTS

I would like to thank my advisor, Mara Breen, not only for her guidance over the course of this thesis, but also for her continual support during my entire four years at Mount Holyoke; from Dr. Seuss to romantic relationships, I have learned and grown extensively as a student and researcher with your help. I would also like to thank Katherine Haydon, who was an integral part of this project, for her dedicated support, and Heather Pon-Barry for her crucial role in this thesis, as well as her continued interest in the project. I also thank Siobhan Norman for her many hours helping with annotations.

Additionally, I would like to thank my family, my rugby team, and my close friends, Cassandra Jonestrask, Kathryn Gill, Evelyn Roberts, Lillian Corman Penzel, Pamela Stürmer, and Zuzanna Przewloka. Your constant encouragement and reassurance made this possible.

Lastly, I thank all of the professors and faculty in the Psychology department. I feel extremely lucky to have been a part of such a welcoming and supportive academic community these last four years.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ABSTRACT

Previous research has demonstrated that speech entrainment, the tendency of conversational partners to match their speech, may be associated with positive social outcomes. The purpose of this study is to investigate speech entrainment in arguably one of the most important contexts: between romantic partners. Through an analysis of acoustic-prosodic features extracted from recordings of twenty-one couples engaged in conversation, I examine (1) the degree to which romantic partners prosodically entrain their speech and (2) what factors contribute to romantic partners' level of prosodic speech entrainment. Using information and behavioral measurements collected from the Couples Communication Project corpus, I investigate whether couples' speech entrainment is dependent on factors such as the length of their relationship, the health of their relationship, and the content of their conversations. Results demonstrate that romantic partners exhibited significant speech entrainment at both a local and global level. No significant differences were found between the degree that couples prosodically entrained and the examined factors. I argue that these findings suggest that speech entrainment between romantic partners is more relevant to momentary interactions, rather than being dependent on general relationship features. I further suggest that this study is limited by the audio recordings and nature of the sample used for analysis, and that extra research is necessary to explore other social factors that may contribute to speech entrainment within romantic relationships.

INTRODUCTION

"Somewhere in the middle of the never-ending noise, there is a pulse, a steady rhythm of a heart that beats, And a million voices blend into a single voice"

–ABBA, *I am the City*

Entrainment can describe any type of matching or synchronization to another being or external rhythm; just as two metronomes placed on a movable base will eventually tick in synchrony with each other, two speakers engaged in a conversation will likely start to speak in a similar manner to each other. The phenomenon of speech entrainment—the tendency to match our speech to our conversational partner(s)—may provide us with significant advantages during daily social interactions (Nenkova, Gravano, & Hirshberg, 2008; Manson, Bryant, Gervais, & Kline, 2013). From forming first impressions, to successfully cooperating with a partner or a group, entraining on speech appears central to our ability to effectively communicate.

Previous work has explored the occurrence of speech entrainment between speakers in a number of social settings, including playing verbal communication computer games (Levitan & Hirschberg, 2011), talking on the phone (Nenkova, Gravano, Hirshberg, 2008), and collaborating during problem solving (Reitter & Moore, 2007; Lubold & Pon-Barry, 2014). Few studies, however, have investigated the occurrence of speech entrainment between non-strangers. The purpose of the current study is to examine the phenomenon of speech entrainment within the context of romantic relationships, and further to use this context as a

means to investigate what factors may contribute to speech entrainment between conversational partners.

Theories of Speech Entrainment

The mechanisms behind our tendency to entrain to one another during dialogue are not exclusive to speech. Accordingly, hypotheses that aim to explain the occurrence of speech entrainment are regularly based in broader theories of behavioral entrainment—including the entrainment of non-verbal actions. One such hypothesis of behavioral entrainment is the *communication accommodation theory* (Giles, Coupland, & Coupland, 1991). Under this theory, people adapt their gestures, speech, and vocal patterns to those around them during social interactions. This theory is also reflected in the *perception-behavior link*, a mechanism proposed by Bargh, Chen, and Burrow (1996), on the basis that people are more likely to use a concept (e.g., perform a certain gesture) if it is already being used in their presence (e.g., others close by performing the gesture), and that therefore, people's behavior is naturally and unconsciously influenced by their perceptions of others behavior around them. That is to say, people are more likely to engage in a particular behavior if they perceive someone else already engaging in it. Evidence for the communication-accommodation theory also comes from research demonstrating that people tend to unconsciously mimic their interaction partners' postures, mannerisms, and body movements—a phenomenon known as the Chameleon Effect (Chartrand & Bargh, 1996).

Another complementary hypothesis of behavioral entrainment is the *coordination-rapport theory* (Tickle-Degen & Rosenthal, 1990). This theory proposes that people's level of non-verbal coordination should relate to their level of rapport, or the degree of harmony and liking between people or groups. In other words, people who are more likely to match each other's gestures and postures are also more likely to enjoy each others company. In accordance with the coordination-rapport theory, Chartrand and Bargh (1996) demonstrated that when participants' postures and movements were matched by a confederate conversational partner, they were prone to smoother interactions and an increased liking of said partner. These results imply that behavioral entrainment may causally lead to increased rapport and better communication. In view of this finding, it has been hypothesized that the tendency to match our social partners' non-verbal and verbal cues likely had an evolutionary survival value in aiding human communication (Lakin, Jefferis, Cheng, & Chartrand, 2003). Given that behavioral entrainment holds such an important communicative value, we can better understand why speech entrainment may be central to successful social interactions.

Speech Entrainment at the Lexical Level

Speech entrainment can be observed at multiple levels. At the lexical and syntactical level (relating to words, grammar, or vocabulary), entrainment can be explored through measurements of word frequency and grammatical/syntactical repetition. These are fairly high (i.e., conceptual) levels of language. At the

phonetic level (relating to the sounds of speech), entrainment can be explored through measurements of the phonetic proximity of words and vowels. This is a lower-level (i.e., perceptual) level of language. The lowest level at which we would expect to find speech entrainment is the level of acoustics. At the acoustic-prosodic level, speech entrainment can be explored using measurements such as pitch, intensity, speaking rate, and voice quality. The latter measurements are those used in the current study. I will describe evidence for speech entrainment at the lexical and phonetic level before focusing on acoustic-prosodic entrainment.

In order to observe speech entrainment at a lexical level, researchers commonly use the Language Style Matching (LSM) algorithm, a method developed by Gonzales, Hancock, and Pennebaker (2010) to calculate the degree to which speakers match their dialogue through an analysis of function words. Function words (e.g., *a, for*, *then, he*) are words that carry little lexical meaning, and instead serve to express grammatical relationships with other words. By measuring the similarity in use of function words, LSM quantifies speakers' non-conscious lexical entrainment independent of the content of their conversation. Research exploring the relationship between LSM and small group social dynamics has demonstrated that a high LSM score between partners, communicating in both face-to-face settings and text-based computer mediated settings, predicts group cohesiveness. Further, this research has demonstrated that a high LSM score within participants communicating in face-to-face settings predicts a more successful performance during search-tasks (Gonzales, Hancock

& Pennebaker, 2010). A related study demonstrated that when three randomly grouped participants conversed for 10 minutes before taking part in an unannounced "prisoner's dilemma" game—a game in which participants are presented with a conflicting scenario and forced to make a decision towards cooperation or self-interest—the pairs of two within the groups with a higher LSM score evaluated each other more positively after the experience (Manson, Bryant, Gervais, & Kline, 2013). LSM score did not, however, predict the likelihood that partners would cooperate during the game.

In addition to LSM, studies of speech entrainment at the lexical level have used other measurements of morphological and grammatical matching. For example, through an analysis of high frequency word use in the Columbia Games Corpus, recordings of randomly paired participants playing verbal communication computer games, and the Switchboard Dialogue corpus, recordings of speakers conversing over the telephone, Nenkova, Gravano, and Hirshberg (2008) demonstrated that a higher degree of word entrainment between speakers was associated with perceived naturalness of dialogue in both copra, and success in the games corpus. These findings demonstrate that speech entrainment between partners is correlated with positive social outcomes in both cooperative and natural settings.

Another technique used for measuring speech entrainment at the lexical level is the Interactive Alignment Model (IAM), a method of speech analysis that measures the extent that conversational partners align situation models during

dialogue (Pickering & Garrod, 2004; Pickering & Garrod, 2006). According to their definition, a situation model is a speaker's multi-dimensional representation of a discussion—including how she represents space, causality, and intentionality. In terms of speech entrainment, alignment of situation models refers to how conversational partners match on dimensions such as grammatical structure (e.g., the card that is blue vs. the blue card), spatial reference frames (e.g., left of the object vs. below the object), and word interpretations. In a study by Reitter and Moore (2007), the researchers used IAM to mark lexical repetition (i.e., word matching) and syntactic repetition (i.e., grammatical structure matching) in order to examine the Human Communication Research Center Map task corpus (HCRC), recordings of participants interacting while completing a route-directing task. In accordance with other studies examining speech entrainment during partner tasks, measurements of lexical and syntactic repetition within the first minutes of dialogue predicted partners' success on the map task.

Phonetic Level Speech Entrainment

While lexical measurements of speech refer to the words of a language, phonetic measurements of speech refer to the sounds or pronunciations of a language that do not differentiate between words. One way to observe speech entrainment at the phonetic level is to measure the phonetic proximity of words, or, the degree that speakers match their pronunciation of identical words. In another speech analysis of the HCRC map task, as measured by the phonetic proximity of words, Pardo (2006) demonstrated that participants matched their

pronunciations of the same words to a greater extent after interacting with each other than before.

Researchers have also investigated phonetic entrainment through the phonetic accommodation of vowels, a measurement of the degree that speakers match their pronunciations of specific vowels. Babel (2012) examined phonetic vowel imitation in a lexical shadowing task in which participants first read a list of words out loud, then listened to either a White or Black model talker read the same words, and lastly identified the words by repeating them out loud again. Results demonstrated that participants phonetically entrained their vowel pronunciations to the model talker, and further that participants differed in how likely they were to entrain based on their racial biases and how attractive they found the talker. Babel argues that these results demonstrate that phonetic entrainment is mediated by subconscious social factors.

At both the lexical and phonetic level, studies exploring speech entrainment have utilized a variety of different measurements; however, they have consistently demonstrated that conversational partners entrain to one another during conversation. Further, this research has reliably shown that higher levels of speech entrainment are associated with positive social factors and interactions (i.e., success on partner computer tasks, naturalness of dialogue, higher degree of rapport or perceived attractiveness). This pattern is also evident in studies investigating speech entrainment at the acoustic-prosodic level.

Acoustic-Prosodic Speech Entrainment

The acoustics of speech are the most basic perceptual elements. These include the perceived pitch of a speaker as high or low, rising or falling; the perceived loudness of a speaker as high or low volume; whether a speaker is talking quickly or slowly; and how a speaker's voice "sounds." This last category is commonly known as voice quality, which will be further explained below. These four categories: Pitch (or frequency), loudness (or intensity), rate, and voice quality comprise *linguistic prosody*. These characteristics are often described as the musical aspects of speech; they are the sounds of the words apart from the sounds that determine the meaning of the word itself. In the current study, I will explore the extent to which speakers entrain to one another on these acoustic features of prosody.

Unlike entrainment on syntactic structures or word choice, prosodic entrainment must be, to an extent, inferred. Assessing lexical entrainment using LSM, for example, requires counting lexical forms. These forms are fairly, though not completely, unambiguous. Prosody, on the other hand, cannot be so directly observed. This disconnect is due to the fact that prosody is a perceptual phenomenon that cannot be directly measured. Therefore, researchers quantify prosody indirectly through acoustic measures that can be directly measured.

To quantify pitch, measurements are taken of a sound's frequency, a measure of how quickly the sound vibrates the air molecules around it. Frequency is measured in Hertz (Hz). Complex sounds like the human voice are comprised

of multiple sound frequencies vibrating simultaneously. Frequencies are periodic if they vibrate in a regular pattern, while aperiodic sounds do not vibrate in a regular pattern. The lowest periodic frequency in a complex sound like speech is that sound's fundamental frequency (F0: "F zero"). To quantify loudness, measurements are taken of a sound's amplitude, or the size of the displacement of air molecules around it. The greater the air displacement, the louder the sound is perceived to be. Measuring intensity requires recording the "size" of the sound as a measure of the amplitude corrected for time and space. To quantify speech rate, a measurement is taken of the number of syllables a speaker produces in a second.

Of all prosodic features, voice quality is the most complex. Voice quality describes the ways in which voices can sound different due to the interactions of other acoustic features. In the current study, I measured four reflections of voice quality: jitter, shimmer, noise-to-harmonics ratio, and harmonics-to-noise ratio. Jitter is a measure of the variability of fundamental frequency over time. Shimmer is a measure of the amplitude variability over time. Noise-to-harmonics ratio and harmonics-to-noise ratio describe the relationship between the proportion of the voice that is harmonic (i.e., has pitch) and the proportion that is aperiodic (i.e., white noise). As these voice qualities can change over time for an individual, they offer an intriguing way to measure entrainment.

Previous research investigating acoustic-prosodic entrainment has demonstrated that, consistent with studies measuring the entrainment of words and pronunciations, the matching of these features across conversational partners

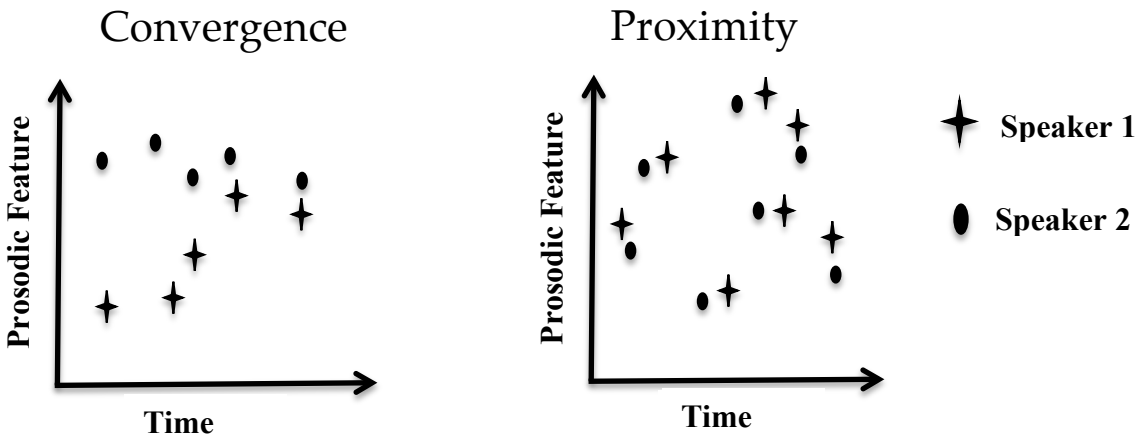
can predict the success of social interactions—and sometimes more consistently than entrainment at the lexical or phonetic level. For example, in a further analysis of the "prisoner's dilemma" game study (discussed above in regards to LSM) at the acoustic-prosodic level, results showed that participants whose dialogue converged on speech rate from the beginning to the end of the conversation were significantly more likely to cooperate during the game (while LSM, on the other hand, was not a significant predictor of cooperation) (Manson et al., 2013). Lubold and Pon-Barry (2014) furthermore demonstrated the social implications of acoustic-prosodic entrainment in the context of collaborative problem solving. To assess the relationship between prosodic speech entrainment and rapport, the researchers carried out an analysis of dialogue from undergraduate students working together in pairs to solve mathematical problems on tablets. Results demonstrated that prosodic entrainment, notably pitch and voice quality, were predictors of perceived rapport between participants.

Levitan and Hirshberg (2011) also analyzed the Columbia Games Corpus at the acoustic-prosodic level, demonstrating that participants exhibited speech entrainment on both a turn-by-turn basis, matching each others prosody at turn exchanges (the time at which one speaker stops speaking and the next one starts talking), and a session basis, matching each others overall prosody. Additionally, this study introduced the framework of observing acoustic-prosodic speech entrainment using measurements of proximity (referred to as "local level" entrainment) and convergence (referred to as "global level" entrainment) that I

adopt in the current study. As measured on a turn-by-turn basis, scores of proximity signify the degree to which partners match their speech prosody at each turn exchange, and scores of convergence signify the degree to which partners' speech prosody becomes more similar throughout a conversation. As illustrated in Figure 1, speakers with high convergence match prosodic features to a greater extent over time, and speakers with high proximity match prosodic features consistently over time.

Figure 1

Visual representations of entrainment measures of convergence (global entrainment) and proximity (local entrainment) (adapted from Levitan & Hirschberg, 2011)



In a more recent study, Levitan et al. (2013) also used the Columbia Games corpus to compare manually labeled social variables with levels of acoustic-prosodic entrainment. Results indicated that higher levels of entrainment were correlated with higher ratings of "giving encouragement" across all gender pairings and higher ratings of "trying to be liked" for female-male and male-male gender pairings. The study also revealed that the highest correlations between level of speech entrainment and social variable ratings occurred in female-male pairings. All of the studies discussed thus far have demonstrated the occurrence of speech entrainment between conversational partners, and suggested that social factors (such as gender and perceived attractiveness of conversational partner) may play a role in determining the extent to which partners entrain their speech. However, this research has focused almost exclusively on interactions between strangers.

Speech Entrainment in Romantic Relationships

Although it is important to observe the dynamics of social exchanges between strangers, some of the most important interactions are arguably those that take place between non-strangers—particularly between romantic partners. Research examining the importance of communication within romantic relationships has demonstrated that effective communication within relationships is central to the satisfaction of romantic partners (Honeycutt & Cantrill, 2014; Meeks, Hendrick, & Hendrick, 1998). Given that speech entrainment has been shown to be associated with more effective and natural dialogue (Nenkova,

Gravano, & Hirshberg, 2008; Reitter & Moore, 2007; Levitan et al., 2013), it is essential to investigate the phenomenon of speech entrainment between romantic partners.

As already noted, few studies have explored speech entrainment between romantic couples; however, speech entrainment within a romantic context has been explored at both the lexical and acoustic-prosodic level. At the lexical level, Ireland et al. (2010) conducted a "speed dating" experiment using the Language Style Matching (LSM) algorithm, in which participants were audio and video recorded during 4-minute dates with the opposite sex. After the interaction, participants rated their date on perceived similarity to themselves, and later reported whether they would be interested in seeing that person again. Results revealed that level of LSM predicted likelihood of mutual romantic interest. Further, a subsequent experiment consisting of an analysis of 10 days worth of instant messages between romantic partners revealed that level of LSM predicted whether or not couples were still dating after 3 months. These findings suggest that dialogue entrainment between partners may have social implications during both initial interactions as well as during later interactions within a romantic relationship.

At the acoustic-prosodic level, Lee et al. (2014) investigated the association between romantic partners' speech entrainment during individual interactions, and the emotional affect of those interactions. Utilizing the Couples Therapy Corpus, a dataset of 134 chronically distressed married couples receiving

therapy, researchers automatically extracted acoustic-prosodic speech features from therapy session conversations. Through a Principal Component Analysis of vocal characteristics, researchers observed correlations between type of affect (negative vs. positive) and the degree to which the couples entrained on their speech during certain interactions. Results demonstrated that level of speech entrainment could predict positive or negative affect 62.56% of time, and that higher levels of speech entrainment were associated with more positive affect. This finding suggests that the extent to which romantic partners entrain their speech can serve to predict the emotional tone of their interaction.

Both of these studies demonstrate that speech entrainment occurs between romantic partners, and that it is commonly associated with positive interactions. However, research thus far has predominantly examined speech entrainment within specific romantic situations: relatively new romantic exchanges (e.g., first dates) and relatively mature romantic relationships (e.g., chronically distressed married couples). Furthermore, studies on the topic of speech entrainment, both between strangers and between romantic partners, have not effectively explored what factors might predict the degree of conversational partners' speech entrainment. Although research has demonstrated an association between speech entrainment and positive interactions, it has not clearly provided a direction of causation for this effect; it is unknown to what extent high levels of speech entrainment lead to positive interactions, or to what extent positive interactions are conducive to higher levels of speech entrainment. Notably, Bock (2012)

examined behavioral entrainment within romantic relationships by measuring the amount that romantic partners rocked in synchrony when sitting in rocking chairs. Results demonstrated that when one partner in the couple was prompted to feel a threat to the relationship, dissatisfied couples started rocking significantly out of synchrony. However, when a relationship threat was introduced to a partner in a satisfied couple, romantic partners demonstrated no difference in the synchronicity of their rocking. These results suggest that behavioral entrainment within romantic relationships may be dependent on factors such as relationship satisfaction and emotional distress. Moreover, these factors may contribute to speech entrainment.

Current Study

The purpose of the current study is to (1) explore the occurrence of speech entrainment within romantic relationships and (2) investigate what social factors affect the degree to which romantic partners entrain their speech. Observing speech entrainment within romantic relationships is an ideal context for examining the social factors that determine the extent to which partners entrain, due to the many quantifiable social factors involved in relationships. Some of these many factors include the length of time that partners have been together and the general health of the partners' relationship. While most previous literature has focused on investigating the social outcomes of speech entrainment (i.e. success on partner tasks, degree of liking of liking between partners), this study aims to explore which factors may lead to and determine the extent of speech entrainment

between romantic partners. In addition to exploring how the length and health of a couple's relationship affects their speech entrainment, I will also explore how the topic of a couple's conversation contributes to speech entrainment.

In order to measure speech entrainment, I carried out an acoustic-prosodic analysis of recordings of couples engaged in (1) a conversation about a conflict within their relationship and (2) a conversation about agreements within their relationship. I measured prosodic speech entrainment on a turn-by-turn basis at both a local level, quantifying how much couples consistently entrained their speech prosody at turn exchanges, and a global level, quantifying how much couples changed their speech prosody to become more similar to one another over the conversation. Using pre-collected survey data and behavioral measurements, I examined the association between the degree that couples entrained their speech and (1) the length of their relationship, (2) the health of their relationship (as quantified by measurements of overall relationship quality, ability to successfully resolve a conflict, and ability to collaborate with one another) and (3) the content of their conversation (e.g., a conversation about a conflict or agreements).

Consistent with past research on speech entrainment between conversational partners, I predicted that romantic partners would entrain their speech prosody at a local level and a global level during both conversations. In line with previous research demonstrating that romantic partners become more similar to one another over time in a variety of ways, including in emotional responses (Anderson, Keltner, & John, 2003), life values (Acitelli, Kennedy, &

Weiner, 2001), and social activities (Price & Vandenberg, 1980), I also predicted that romantic partners who have been in a relationship for longer would demonstrate a higher degree of speech entrainment. Additionally, in accordance with research suggesting that romantic partners exhibit more speech entrainment during interactions characterized by positive affect (Lee et al., 2014), I expected that couples in healthier relationships would exhibit higher levels of speech entrainment due to an increased number of positive interactions. Further, I predicted that couples would demonstrate more prosodic entrainment when conversing about agreements (likely with higher levels of positive affect), than when conversing about a conflict (likely with higher levels of negative affect).

METHOD

Participants

The current study used the data (audio recordings and behavioral measurements) of 21 couples (42 total participants; 18 heterosexual couples, three same-sex couples) randomly drawn from the Couples Communication Project corpus (Haydon, Jonestrask, Guhn-Knight, & Salvatore, under review). All participants from the corpus were recruited via flyers and online advertisements from around the Connecticut River Valley Area in Massachusetts, and met the criteria of being (1) over 18 years of age and (2) in their current relationship for at least one year. The 42 participants whose data was used in the current study ranged in age from 19 years old to 43 years old, with an average age of 26 (SD = 4.7), and were in relationships that ranged in length from one year to six and a half years, with an average length of 2.9 years (SD = 1.9),

Materials

Couples Communication Project Corpus

The Couples Communication Project Corpus (Haydon, et al., under review) is an existing dataset of one hundred couples who completed and participated in (1) an online survey of demographic information and family relationship history, (2) a discussion task in which they conversed about mutual disagreements and agreements within their relationship, and (3) an interview session in which they answered questions about their relationship history.

Additionally, couples were contacted one year after the initial visit for a follow up survey on whether or not they were still romantically involved.

The current study focuses primarily on data from the discussion task, yet additionally utilizes information from the interview session and initial online survey. Information drawn from the online survey and used in this study included the length of each couples relationship. Information drawn from the interview session included a dyadic score of *interview collaboration*, given to each couple as a measurement of how much they collaborated in order to tell their relationship history. The interview session consisted of a joint interview in which couples were instructed to tell the history of their relationship together. Coders later watched the interview and provided each couple with a collaboration score (rated on a 1-5 scale) [Full scale in Appendix B]. Couples who received higher scores on this measure appeared to work together to tell their relationship history, build off of one another's stories, and view the relationship in a similar manner. Couples who received lower scores on this measure appeared to not work together when telling their relationship history, and tended to invalidate or disagree with each other's stories. Coders exhibited high inter-rater reliability of interview collaboration based on 23% the sample, with an intraclass correlation of .87.

The discussion task was comprised of the Markman-Cox Procedure (Cox, 1991), which consists of a conflict section and recovery section. During the conflict section (10 minutes) partners were instructed to choose a topic or part of their relationship (e.g., communication, in-laws, recreation, religion, etc.) that

they disagreed on and then try to resolve the conflict. During the recovery section of the discussion task (4 to 5 minutes) partners were instructed to choose topics or parts of their relationship that they mutually agreed on and to discuss them. Multiple trained coders watched the audio-video recordings of the couples' conversations and provided quantitative behavioral measurements of both individual behavior, in which each participant received their own score, and dyadic functioning, in which each couple received a joint score.

This study focuses on the two measurements of dyadic functioning derived from the discussion task. That is, measurements in which a single score was given to each couple as unit. These measurements consisted of *dyadic conflict resolution* and *overall quality* (both rated on a 1-7 scale) [Full coding scales in Appendix A]. The measurement of dyadic conflict resolution reflected how well the couple worked together to resolve the conflict (during the conflict section) in a mutually satisfying way. Couples received higher scores on this measure if they demonstrated that they were able to work towards a common goal cooperatively, and if they appeared mutually satisfied with the process. Conversely, couples received lower scores if they were insensitive to each other's feelings, and if only one partner appeared satisfied with the resolution. Coders exhibited high inter-rater reliability of dyadic conflict resolution based on 23% of the sample, with an intraclass correlation of .90.

The measurement of overall quality reflected the general sense of quality of the relationship for both partners. Couples received higher scores on this

measure if they appeared to show mutual caring, trust, emotional closeness, and sensitivity and enjoyment of one another, while couples received lower scores if they appeared to lack these qualities and instead show negative characteristics such as victimization, chronic intense conflict, or rigidity of roles. Coders exhibited high inter-rater reliability of overall quality based on 23% sample, with an intraclass correlation of .94. In the current study, I quantified health of relationship using scores of the two dyadic measurements from the discussion task (dyadic conflict resolution and overall relationship quality), and the dyadic measurement from the interview session (interview collaboration) (Table 1). The twenty-one couples analyzed in the current study varied in each of these dyadic measurements as shown in Figure 2.

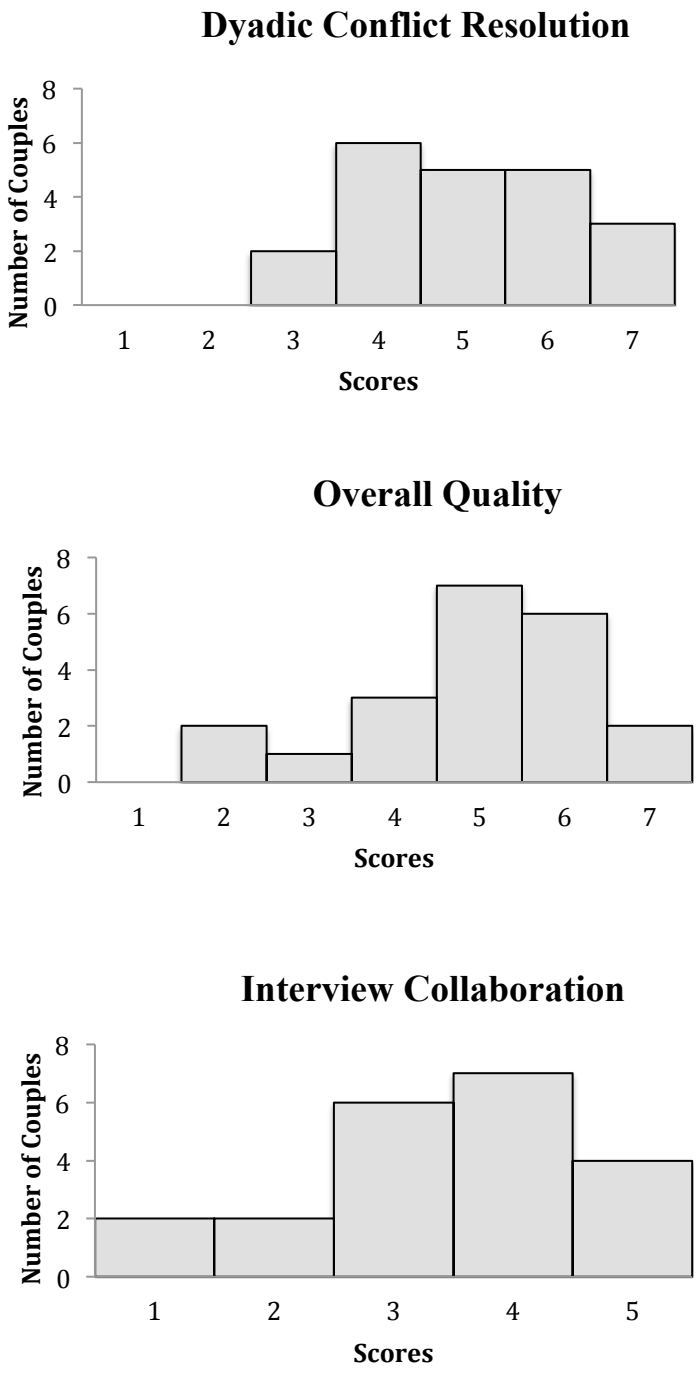
Table 1

Couples Communication Project corpus procedure and behavioral measurements used to examine social factors

Social Factor	Procedure	Measurement	
Length of Relationship	Online Survey	Length of relationship (in years)	
Health of Relationship	Discussion Task	Overall Relationship Quality (1-7)	Dyadic Conflict Resolution (1-7)
	Interview Session	Interview Collaboration (1-5)	
Content of Conversation	Discussion Task	Audio of Conflict Section	Audio of Recovery Section

Figure 2

Variance in couples' dyadic scores of conflict resolution, overall quality, and interview collaboration used to quantify health of relationship



Procedure

Audio Preparation

The audio tracks of the couples' interactions were exported from the video recordings at a sample frequency of 48kHz as stereo wav files and imported into Praat (Boersma & Weenink, 2012), a specialized software program for speech analysis. Acoustic measures of prosody were extracted from each conversation using the following method:

Identifying Inter-Pausal Units

In order to measure acoustic-prosodic speech entrainment on a turn-by-turn basis, it is necessary to compare the last segment of a speaker's turn with the first segment of his or her partner's turn. These segments are referred to as inter-pausal units (IPUs). The process of identifying IPUs began with marking speaker backchannels, speaker overlap, and speaker turns. Speaker *backchannels* were defined as speech produced purely for phatic communication, i.e., lacking conversational content, and were usually signs of understanding, affirmation, or disagreement with a partner (e.g., "okay," "mm-hmm," "yea," "uh-uh"). Speaker *overlap* was defined as any simultaneous speech by both partners. Speaker *turns* were defined as continuous speech by a single speaker, interrupted only by overlap or backchannels that lasted less than 400ms. This length of time was empirically derived from the annotation of 10 couples as the approximate time that it took for speakers to adjust to the fact that their partner was talking.

Inter-pausal units were extracted from speaker turns, and identified as pause-free segments of speech from a single speaker that were (a) separated from one another by at least 50ms (as defined by Levitan & Hirshberg, 2011), and (b) not interrupted by instances of backchannels or overlap. Both backchannels and overlap were not considered IPUs, and therefore were excluded from the final speech analysis. In the atypical absence of a silence or interruption during a speaker's turn, IPUs were also marked at clear sentence or phrase boundaries, as indicated by characteristic pitch movements or syllable lengthening. This process helped identify IPUs that were too long to reliably exhibit local entrainment. In order to account for IPUs that were too short to have reliable prosodic characteristics, IPUs that were less than 50 ms long were excluded from analysis. Additionally, IPUs that were one or two syllables were only identified if they were more than 90 ms long, so as to get an accurate measurement of speech rate.

Only the initial and final IPUs of each speaker's turn were used for analysis. In the case that a turn contained only one IPU, the IPU was marked as both final and initial. During analysis, the prosodic characteristics of all adjacent final and initial IPUs from alternating speakers were compared. To illustrate, in the speech example below, the initial IPUs are bold and the final IPUs are underlined. In this example the prosodic characteristics of Speaker One's final IPU ("but when I looked outside I decided to go back to sleep") would be compared with the prosodic characteristics of Speaker Two's initial IPU ("That sounds like my morning too"):

Speaker 1: **"This morning I woke up early** [pause] but when I looked outside I decided to go back to sleep"

Speaker 2: **"That sounds like my morning too** [pause] except for [pause] well I decided to make breakfast"

Speaker 1: **"Well I think you made a good choice**. [clear sentence boundary] Do you know what time you will get home tonight?"

Speaker 2: **"Not before nine o' clock"**

Extracting Acoustic Measures

I extracted the following prosodic measures of pitch, amplitude, speech rate, and voice quality from each IPU (Table 2):

1) Pitch: Using Praat's scripting tool, I extracted the mean and maximum fundamental frequency (F0) of each initial and final IPU.

2) Intensity: Using Praat's scripting tool, I extracted the mean and maximum intensity of each initial and final IPU.

3) Speech Rate: Using Praat's scripting tools, syllable boundaries in each wav file were automatically identified whenever the amplitude of the speaker's voice changed by 40 decibels (dB) or more. I then extracted information about the timing of each syllable in order to determine the number of syllables in each initial and final IPU.

4) Voice Quality: Using Praat's scripting tool, I extracted measurements of jitter, shimmer, noise-to-harmonics ratio, and harmonics-to-noise ratio from each initial

and final IPU. Jitter was measured by calculating the variation in fundamental frequency across each IPU. Shimmer was measured by calculating the variation in intensity across each IPU. Noise-to-harmonics ratio and harmonics-to-noise ratio were calculated by measuring the proportion of the speech within each IPU that was harmonic and the proportion that was aperiodic

Table 2

Acoustic-prosodic features extracted from each IPU used for entrainment analysis

Acoustic Feature	Units	Description
Mean F0	Hz	Mean F0 for the entire IPU
Maximum F0	Hz	Maximum F0 across the entire IPU
Mean intensity	dB	Mean intensity of the entire IPU
Maximum intensity	dB	Maximum IPU across the entire IPU
Speech Rate	syll/sec	Number of syllables in IPU divided by number of seconds in IPU
Jitter	%	Variability of F0 across entire IPU
Shimmer	%	Variability of intensity across entire IPU
Noise-to-Harmonics	:	Proportion of aperiodic to speech to harmonic speech across entire IPU
Harmonics-to-Noise	:	Proportion of harmonic speech to aperiodic speech across entire IPU

Measurements of Acoustic-Prosodic Entrainment

In the current study I quantify speech entrainment on a turn-by-turn basis using measurements of (1) proximity (or local level entrainment): the similarity of partners' speech at turn exchanges, and (2) convergence (or global level entrainment): the tendency of partners' speech to become more similar over turn exchanges. For each measure, the acoustic-prosodic features of pitch, intensity, speech rate, and voice quality were compared at adjacent IPUs and used to calculate entrainment.

Local measurement of Proximity:

Scores of proximity on a turn-by-turn basis indicate how consistently similar a couple's speech prosody was across turn exchanges. Proximity was measured by calculating the absolute value of the prosodic differences between speakers' adjacent IPU's ("partner distance"), and comparing that value to the absolute value of the prosodic differences between a target IPU and five other random non-adjacent IPUs ("other distance") (Figure 3). Individual entrainment scores for each couple were computed by subtracting the "other distance" from the "partner distance," meaning that negative scores indicated higher levels of entrainment. Entrainment as measured by proximity was inferred if the absolute value for "partner distance" was significantly less than the absolute value for the "other distance." For the purpose of presenting results, scores will be reversed so that higher scores indicate higher entrainment.

Figure 3

Equations used to calculate proximity on a turn-by-turn basis (adapted from Levitan & Hirschberg, 2011)

$$\boldsymbol{partner\ distance} = \boldsymbol{|IPU(target) - IPU(partner)|}$$

$$\boldsymbol{other\ distance} = \frac{\boldsymbol{\sum_{i=1}^{5} |IPU(target) - IPU(i)|}}{\boldsymbol{5}}$$

Global measurement of Convergence:

Scores of convergence at on a turn-by-turn basis indicate the tendency of partners' speech prosody to become more similar to each other across turn exchanges over the length of a conversation. Entrainment scores were calculated by computing the correlation between the absolute value of prosodic differences of adjacent IPU's and the IPU number (used as a measurement of time). Entrainment as measured by convergence was inferred if the correlation was significant.

RESULTS

Speech entrainment at a local and global level, as measured on a turn-by-turn basis, was calculated across all couples for both the conflict section and recovery section of the discussion task. In order to measure local level entrainment (proximity), I conducted paired t-tests between the acoustic features of "partner difference" (adjacent IPUs) and "other difference" (five random non-adjacent IPUs). In order to measure global level entrainment (convergence), I ran a correlation between the acoustic features of adjacent IPUs and the IPU number.

Results demonstrated that romantic partners exhibited significant speech entrainment at both a local level and a global level. At the local level, speakers significantly matched their romantic partners' speech prosody on features of intensity and voice quality; during the conflict section couples exhibited significant local entrainment on mean intensity ($p < .001$), maximum intensity ($p < .001$), and shimmer ($p < .05$) (Table 3), and during the recovery section couples exhibited significant local entrainment on mean intensity ($p < .001$), maximum intensity ($p < .001$), jitter ($p < .05$), shimmer ($p < .05$), and Noise-to-Harmonics ratio ($p < .001$) (Table 4). At the global level, speakers' speech prosody over the conversation became significantly more similar to their romantic partners on features of pitch, intensity, voice quality, and speech rate; during the conflict section couples exhibited significant global entrainment on mean pitch ($p < .01$), mean intensity ($p < .001$), maximum intensity ($p < .05$), and speech rate ($p < .05$) (Table 5), and during the recovery section couples exhibited significant global

entrainment on mean pitch ($p < .001$), maximum intensity ($p < .001$), and Noise-to-Harmonics ratio ($p < .05$) (Table 6). Speech rate was unable to be extracted for measurements of local entrainment and measurements of global entrainment for the recovery section due to missing data.

Table 3

*Measurements of Proximity during **conflict section** of discussion task: t-test of partner difference vs. other difference*

Feature	t	df	p-value	Sig
1 Mean Pitch	-1.17	20	.25	
2 Max Pitch	.83	20	.42	
3 Mean Intensity	-3.27	20	.00	**
4 Max Intensity	-4.63	20	.00	**
5 Jitter	-1.52	20	.14	
6 Shimmer	-2.18	20	.04	*
7 Noise-to-Harmonics	-1.40	20	.17	
8 Harmonics-to-Noise	-1.61	20	.12	

Table 4

*Measurements of Proximity during **recovery section*** of *the discussion task: t-test of partner difference vs. other difference*

Feature	t	df	p-value	Sig
1 Mean Pitch	-.26	20	.79	
2 Max Pitch	-.00	20	.10	
3 Mean Intensity	-3.97	20	.00	**
4 Max Intensity	-3.23	20	.00	**
5 Jitter	-2.65	20	.02	*
6 Shimmer	-1.99	20	.04	*
7 Noise-to-Harmonics	-3.05	20	.01	**
8 Harmonics-to-Noise	-2.03	20	.06	.

Table 5

*Measurements of Convergence during **conflict section** of discussion task: correlation between partner difference vs. IPU number*

Feature	t	df	p-value	corr	Sig
1 Mean Pitch	2.71	1353	.01	.07	**
2 Max Pitch	-1.00	1353	.32	-.03	
3 Mean Intensity	-3.21	1353	.00	-.09	**
4 Max Intensity	-2.30	1353	.02	-.63	*
5 Jitter	1.68	1353	.09	.05	.
6 Shimmer	-0.36	1353	.71	-.01	
7 Noise-to-Harmonics	.20	1353	.84	.01	
8 Harmonics-to-Noise	-1.53	1353	.13	-.04	
9 Speech Rate	2.51	1353	.01	.07	*

Table 6

*Measurements of Convergence during **recovery section** of discussion task: correlation between partner difference vs. IPU number*

Feature	t	df	p-value	corr	Sig
1 Mean Pitch	4.10	632	.00	.16	**
2 Max Pitch	1.06	632	.29	.02	
3 Mean Intensity	-1.46	632	.14	-.06	
4 Max Intensity	-2.86	632	.00	-.11	**
5 Jitter	-3.19	632	.75	-.01	
6 Shimmer	-.73	632	.46	-.03	
7 Noise-to-Harmonics	-3.41	632	.73	-.01	*
8 Harmonics-to-Noise	-2.19	632	.03	-.09	

In order to investigate whether there was a relationship between the degree that couples entrained their speech and the length and/or health of their relationship, I first conducted a Principal Components Analysis (PCA) of the eight extracted acoustic-prosodic features (excluding speech rate) as measured by both proximity and convergence. After examining the scree plot and eigenvalues, I determined that a two-component solution best accounted for the data. Components were derived by retaining features that loaded above 0.6 and did not cross-load other factors above 0.4 (Haydon, Roisman, & Burt, 2012). As measured by proximity, the first component reflected the prosodic features of voice quality (Voice Quality factor, eigenvalue = 3.06) and consisted of measurements of shimmer, noise-to-harmonics ratio, and harmonics-to-noise ratio, while the second component reflected the prosodic features of pitch and

loudness (Pitch Loudness factor, eigenvalue = 1.70) and consisted of measurements of mean pitch and maximum intensity (Table 7). As measured by convergence, the first component reflected the prosodic features of pitch and loudness (Pitch Loudness factor, eigenvalue = 2.43) and consisted of measurements of maximum pitch, maximum intensity and mean intensity, while the second component reflected the prosodic features of voice quality (Voice Quality factor, 0value = 1.78) and consisted of measurements of noise-to-harmonics ratio and harmonics-to-noise ratio (Table 8).

Table 7

Component matrix of acoustic-prosodic features for discussion task as measured by proximity

	Component	
Speech Feature	**1**	**2**
1 Mean Pitch	-.183	**.775**
2 Max Pitch	.482	.172
3 Mean Intensity	.560	.561
4 Max Intensity	.318	**.743**
5 Jitter	.488	.250
6 Shimmer	**.754**	.102
8 Noise-to-Harmonics	**.878**	.001
9 Harmonics-to-Noise	**.847**	-.422

Note: **Factor 1 is Voice Quality (eigenvalue = 3.06), Factor 2 is Pitch Loudness (eigenvalue = 1.70). Bold factor loadings reflect items retained in final scales.**

Table 8

Component matrix of acoustic-prosodic features for discussion task as measured by convergence

	Component	
Speech Feature	**1**	**2**
1 Mean Pitch	.336	.009
2 Max Pitch	**.653**	.028
3 Mean Intensity	**.701**	.432
4 Max Intensity	**.731**	.366
5 Jitter	-.802	.325
6 Shimmer	.033	.123
8 Noise-to-Harmonics	-.171	**.850**
9 Harmonics-to-Noise	.115	**.889**

Note: **Factor 1 is Pitch Loudness (eigenvalue = 2.43), Factor 2 is Voice Quality (eigenvalue = 1.78). Bold factor loadings reflect items retained in final scales**

Using these PCA derived factors of Voice Quality and Pitch Loudness, I ran a linear regression in order to examine the relationship between each couple's dyadic level of speech entrainment (at both a local level and global level), and the length of their relationship (Table 9). I found no significant relationship between level of speech entrainment, as measured by the Voice Quality factor or the Pitch Loudness factor, and length of relationship.

Table 9

Results of linear regression model predicting relationship length from the PCA derived acoustic variables

	Length of Relationship			
	b	*SE*	*T*	*P*
	Proximity			
PCA Voice Quality	3.27	3.23	1.01	.33
PCA Pitch Loudness	.27	.29	.94	.36
	Convergence			
PCA Pitch Loudness	-.50	1.32	-.37	.71
PCA Voice Quality	2.64	1.545	1.71	.11

In order to quantify relationship health, I examined behavioral measurements of (1) overall quality, (2) dyadic conflict resolution, and (3) interview collaboration. Using the PCA derived factors, I ran a linear regression to examine the association between these behavioral measurements and each couples dyadic level of speech entrainment (Table 10). Results demonstrated no significant relationship between level of speech entrainment and the examined behavioral measurements. The relationship between couples' scores of dyadic conflict resolution and level of speech entrainment as measured by the Pitch Loudness factor approached, but did not reach, statistical significance at both a local level (p= .076) and a global level (p= .069).

Table 10

Results of linear regression model predicting relationship health (as measured by dyadic conflict resolution, dyadic overall quality, and interview collaboration) from the PCA derived acoustic variables

	Dyadic Observed Quality			
	b	*SE*	*T*	*P*
		Proximity		
PCA Voice Quality	-.24	2.59	.091	.93
PCA Pitch Loudness	.39	.23	-1.69	.11
		Convergence		
PCA Pitch Loudness	1.72	1.12	1.53	.14
PCA Voice Quality	-1.25	1.31	-.95	.352

	Dyadic Conflict Resolution			
	b	*SE*	*T*	*P*
		Proximity		
PCA Voice Quality	-.33	2.34	.143	.89
PCA Pitch Loudness	.39	.207	-1.88	**.08**
		Convergence		
PCA Pitch Loudness	1.91	.99	-1.94	**.07**
PCA Voice Quality	-1.57	1.15	-1.36	.19

	Interview Collaboration			
	b	*SE*	*T*	*P*
		Proximity		
PCA Voice Quality	-2.68	2.04	1.32	.21
PCA Pitch Loudness	-.10	.18	.55	.59
		Convergence		
PCA Pitch Loudness	1.44	.83	1.73	.10
PCA Voice Quality	-1.18	.97	-1.21	.24

So as to investigate the relationship between the degree that couples entrained their speech prosody and the content of their conversation, I conducted paired t-tests between the conflict section and recovery section of the discussion task across all couples. No significant differences were found between the two conversations at a local level (as measured by proximity) (Table 11), or at a global level (as measured by convergence) (Table 12).

Table 11

T-tests of conflict section vs. recovery section entrainment for discussion task as measured by proximity

Feature	t	df	p-value	Sig
1 Mean Pitch	.50	20	.61	
2 Max Pitch	-.29	20	.76	
3 Mean Intensity	-1.78	20	.08	.
4 Max Intensity	-.77	20	.44	
5 Jitter	-.65	20	.51	
6 Shimmer	-.12	20	.90	
7 Noise-to-Harmonics	-.52	20	.60	
8 Harmonics-to-Noise	-.15	20	.87	

Table 12

T-tests of conflict section vs. recovery section entrainment for discussion task as measured by convergence

Feature	t	df	p-value	Sig
1 Mean Pitch	-.43	20	.67	
2 Max Pitch	.07	20	.94	
3 Mean Intensity	.08	20	.93	
4 Max Intensity	-.75	20	.45	
5 Jitter	-.29	20	.77	
6 Shimmer	-.22	20	.82	
7 Noise-to-Harmonics	.11	20	.91	
8 Harmonics-to-Noise	-.16	20	.85	

DISCUSSION

The current study was designed to examine acoustic-prosodic speech entrainment between romantic partners, and to investigate what social factors best predict the extent of their prosodic entrainment. Results demonstrated that romantic partners engaged in conversation entrained acoustic features of their speech at both a local level and global level. This outcome is consistent with my hypothesis, as well as previous research on acoustic-prosodic speech entrainment between conversational partners (Levitan & Hirschberg, 2011; Manson et al., 2013; Lubold & Pon-Barry, 2014, Lee et al., 2014). These results reveal that during conversation romantic partners tend to (1) consistently talk with similar prosody, matching the prosodic features of their significant other during turn exchanges, and (2) increase the similarity of their prosody over the length of a conversation, matching each other's prosodic features to a greater extent at the end of a conversation than the beginning. Results also demonstrated that level of prosodic speech entrainment between romantic partners was not related to factors such as the length of their relationship, the health of their relationship, or the content of their conversation. I argue that there are multiple explanations for these findings: (1) there is no relationship between the examined factors and level of speech entrainment; (2) the audio recordings of the conversations used in the current study prevented accurate measurements of speech entrainment; (3) the nature of the sample of the current study was too limited to reveal a systematic relationship.

Speech Entrainment and Length of Relationship

Results did not show any significant associations between the extent to which romantic partners entrained their speech and the length of their relationship. This finding was contrary to previous research suggesting that romantic partners tend to become more similar over time in a number of aspects of their emotional and social lives (Anderson, Keltner, & John, 2003; Acitelli, Kennedy, & Weiner, 2001; Price & Vandenberg, 1980). This result, however, is not unexpected given that previous research has not explored an effect of relationship length on speech entrainment. Rather, studies on speech entrainment have consistently shown that partners, including strangers, regularly exhibit high levels of speech entrainment within their first conversation (Levitan & Hirshberg, 2011; Nenkova, Gravano, & Hirshberg, 2008)—and further, that conversational partners exhibit increased entrainment over the course of the conversation (in the current study this happens in a conversation span as short as five minutes). Given the speed that conversational partners are able to entrain their speech, it seems likely that romantic partners who have been together for more than a year already display a certain level of prosodic entrainment as a result of their time together (signifying a ceiling effect). While individual couples still significantly vary in the degree that they entrain their speech to one another, even after a year together (Ireland et al., 2012), I argue that the effect of length of relationship on acoustic-prosodic entrainment likely plays a role at a much earlier stage. Therefore, in order to observe an effect of relationship length it may be necessary to

longitudinally measure the speech entrainment of couples within the first few months, or even weeks, of their relationship.

Additionally, this finding may be a result of the relatively small variation in relationship length within the couples in the current study. Couples included in the current analysis ranged in length of relationship from one year to six and a half years (SD = 1.9). It is likely that the length of a relationship has a long-term effect on romantic partners' prosodic speech entrainment, yet that these effects can only be detected in a broader sample. For example, one possibility is that couples who have been together for more than twenty years exhibit higher levels of speech entrainment than couples who have been together from one to five years. However, that this difference is harder to detect in relatively younger couples because, after an immediate and rapid increase in entrainment within the first few weeks or months, level of speech entrainment increases at a much slower rate.

Speech Entrainment and Health of Relationship

Health of relationship did not have a significant effect on the extent to which couples entrained their speech as measured by the overall quality of romantic partners' relationships, the ability of the romantic partners to resolve conflicts, or the ability of romantic partners to collaborate on telling their relationship histories. Although I predicted that couples in healthier relationships would exhibit higher levels of prosodic entrainment, previous studies have also failed to find this effect. In an examination of entrainment within couples as

measured by chair rocking synchrony, Bock (2012) found an interaction effect between the level of couples relationship satisfaction and their sustained entrainment after a relationship threat. However, the study showed no significant difference in baseline rocking entrainment between satisfied and dissatisfied couples. Similarly, Ireland et al. (2012) found that while speech entrainment as measured by LSM was able to predict relationship stability at three months, level of entrainment was unrelated to self-reported relationship quality and satisfaction. The results of the current study are consistent with these findings, demonstrating that couples in relatively healthy and relatively unhealthy relationships do not differ in the extent to which they prosodically entrain on speech.

Conversely, Lee et al. (2014) demonstrated that romantic partners exhibit higher levels of speech entrainment when involved in interactions characterized by positive affect. In view of the findings of the current study in conjunction with this previous research, I suggest that speech entrainment is less dependent on large-scale factors such as overall relationship health and partner satisfaction, and more dependent on smaller scale factors such as the tone and success of an particular exchange.

Notably, the measurement of *dyadic conflict resolution*, used in the current study to quantify relationship health by recording the extent to which couples successfully worked together to resolve a conflict, may provide the best measurement of the success of an individual interaction. Although high scores of dyadic conflict resolution do not necessarily suggest positive affect, they do

suggest a relatively effective interaction that likely included cooperation, understanding, respect, and sensitivity to one another during the conflict section of the discussion task [Full Coding Scale in Appendix A]. Additionally, couples who scored higher on measurements of dyadic conflict resolution were likely more relaxed, and therefore displayed more positive affect, during the recovery section of the discussion task. Although the regression between couples' dyadic conflict resolution scores and degree of speech entrainment was not significant, results suggested an emerging relationship such that couples with higher levels of prosodic speech entrainment as measured by the PCA derived Pitch Loudness factor scored higher on the measurement of dyadic conflict resolution at both a local level ($p = .076$) and global level ($p = .069$). I argue that this emerging trend, in combination with the study by Lee et al. (2014), provides further evidence that speech entrainment is most dependent on smaller scale factors such as the affect and effectiveness of a current exchange between speakers.

Speech Entrainment and Content of Conversation

Results suggesting that the content of couple's conversations did not have an effect on their prosodic speech entrainment are also inconsistent with the findings of Lee et al. (2014). Given that couples tend to entrain more during positive interactions, I expected that couples discussing agreements would exhibit greater positive affect, and therefore more entrainment, than couples discussing a conflict. However, the disparity between the current results and prior research may be because this study specifically looked at conversation *content* rather than

conversation *affect.* That is, while Lee et al. (2014) used specific measurements of "global positive" and "global negative" to measure the affect and tone of couples interactions, the current study simply measured the presence of an agreement or conflict during each discussion. Provided that couples varied in the degree to which they positively or negatively approached the conversation about a conflict and the conversation about agreements, this comparison of conversation content does not consistently predict the tone or affect these interactions. Rather, in order to quantify the affect of these interactions, dyadic measurements of positive and negative affect from each conversation would need to be examined in relation to the extent to which partners prosodically entrain their speech.

Further, the lack of difference in speech entrainment between the conflict section and recovery section may be an effect of the procedure of the discussion task; while couples conversed about a conflict for ten minutes, they only conversed about agreements for five minutes. In order to take a more accurate measurement of the effect of conversation content on degree of speech entrainment, it may be necessary for romantic partners to have the same amount of time to talk about each conversation topic. This is particularly important given that the current study demonstrated significant global entrainment across couples; because couples continued to entrain over the length of each conversation, the addition of five minutes of talking may have significant consequences on the overall degree of prosodic entrainment.

Study Limitations

As is often the case when utilizing a corpus of speech previously developed for another purpose, the Couples Communication corpus was not created for the purpose of acoustic analysis, and therefore presented limitations in regards to measuring acoustic-prosodic entrainment. One such limitation was the way in which the audio was recorded. In contrast to methods of previous studies investigating acoustic-prosodic entrainment, which typically record conversational partners on two separate audio channels (Lubold & Pon-Barry, 2014; Levitan & Hirshberg, 2011), the current study used speech extracted from a single-channel audio recording of both partners. This meant that in order to examine acoustic-prosodic speech features of each partner individually, speakers' turns had to be marked and extracted by hand. Further, a single-channel audio recording meant that any time the audio was interrupted by the non-speaking partner (i.e. laughing, coughing, sneezing, grumbling), it could not be used for analysis. Most importantly, using these audio recordings of the conversations didn't allow for the analysis of any partner overlap; even if partners were talking simultaneously with high levels of prosodic entrainment, their speech could not be used because the prosodic features of each speaker could not be separated.

Additionally, the single-channel recording also presented a limitation in the case of backchannels. To illustrate, when the non-primarily talking partner expressed affirmation at their partners' dialogue (e.g. "mmhmmm"), neither

partners' segments of speech could be used for analysis due to simultaneous audio and therefore non-extractable acoustic features. Although these disruptions did not occur frequently in most couples' conversations, it is important to take them into account in light of previous research demonstrating that higher degrees of speech entrainment are associated with more overlap and fewer interruptions between speakers (Nenkova, Gravano, & Hirshberg 2008). As defined by Nenkova, Gravano and Hirshberg, *overlap* refers to when Speaker Two starts her turn before Speaker One is done, and when Speaker One still completes her turn, while *interruption* refers to when Speaker Two starts her turn before Speaker One is done, but when Speaker One does not complete her utterance. In the current study, however, these are not differentiated. Rather, any simultaneous talking between speakers could not be utilized. It is necessary that future research examining prosodic speech entrainment within romantic relationships use a separate two-channel audio recorder in order to examine differences in entrainment during overlaps and interruptions, and provide a better measurement of overall level of acoustic-prosodic entrainment.

This study was also limited by the relatively small sample size of couples whose conversations were used for analysis. The process of preparing the conversation audio for an analysis of speech entrainment took approximately two to three hours for each couple. In contrast to the current study's analysis of twenty-one couples, previous research demonstrating a significant relationship between the degree that couples entrained their speech and the emotional affect of

their interactions, sampled over one-hundred couples (Lee et al., 2014). It is likely that the findings of the current study do not accurately reflect the relationship between the examined social factors and level of speech entrainment due to a lack of data from a small number of couples.

Additionally, given that speech entrainment is dependent, at least in part, on social factors, a larger sample size would provide more variation across these factors. Although the couples whose data was utilized in the current study ranged considerably in the health of their relationship, their scores were generally towards the higher end of the spectrum; on measurements of dyadic conflict resolution and overall quality very few or no couples scored below a three (Figure 2). In order to find a relationship between the health of a couple's relationship and the level of their speech entrainment it may be necessary to include an analysis of speech from lower scoring, and therefore less healthy, couples.

Future Directions

The current study has many implications for future directions of research. In particular, I suggest that studies moving forward should (1) account for the number of turns and overlaps within conversations between romantic partners, (2) examine the acoustic-prosodic entrainment of backchannels within these conversations, (3) investigate acoustic-prosodic speech entrainment within these conversations on a session basis and (4) examine speech entrainment between romantic partners with a combination of lexical, phonetic, and acoustic-prosodic entrainment measures.

One of the most noticeable variations between couples' conversations was the pattern of turn taking. To clarify, turn exchanges refer to the point in time that one partner stops talking and the next partner starts talking. Turn exchanges are necessary in order to measure prosodic entrainment on a turn-by-turn basis, as it is also the point in time where the prosodic features of adjacent final and initial IPUs are compared. In an examination of the ten minute long conflict section of the discussion task, I found that couples took as few as 11 turn exchanges and as many 114 turn exchanges. During the conversations, some couples took turns speaking evenly and at regular intervals (i.e., both partners sharing ideas and responding to the ideas of their partner in a back and forth manner). Other couples, however, took turns speaking very unevenly and irregularly (i.e., one partner sharing and talking significantly more than the other partner). Within couples that talked in an even back and forth manner, conversations included many turn exchanges. Therefore, in these conversations I was able to extract and analyze many adjacent IPUs for analysis. However, within couples that demonstrated uneven and irregular turn exchanges, I was able to extract significantly fewer adjacent IPUs for analysis. Given that level of speech entrainment is dependent on the prosodic data extracted from these adjacent IPUs, much less data was available to determine entrainment for couples that spoke with irregular turn exchanges.

Another factor that contributed to variability in turn number was difference in overlap (as defined by any simultaneous talking from both speakers) across

couples. Specifically, the conversations from couples that tended to overlap each other contained fewer IPUs, and therefore provided a sparser dataset to measure entrainment. In order to account for these differences, it is necessary that future research a) use a two-channel audio recorder to measure entrainment during overlap and b) account for number of turn exchanges by using the same number of IPUs from each couple to measure entrainment.

On the other hand, variability in turn exchanges between partners, due to talking patterns and overlap, could also be a direction for future research. Using measurements of turn exchanges, future studies may examine what social factors are related to higher and lower numbers of turn taking. Research could also examine the relationship between social factors and the number of overlap and interruptions (as defined by Nenkova, Gravano and Hirshberg, 2008) that occur between partners. For example, studies can investigate whether couples in a healthier relationship, or involved in a more positive interaction, talk with more regular or irregular turn exchanges, or more overlap or interruptions.

It is also important that future research investigate the prosodic entrainment of backchannels within these conversations. To recapitulate, backchannels are purely phatic expressions, commonly signs of affirmation or agreement, from the non-predominantly talking speaker (e.g. “mhmm” “right” “nahh”). Previous studies have examined conversational partners’ entrainment of backchannels, demonstrating that speakers’ backchannels become prosodically more similar throughout a conversation (Levitan, Gravano, & Hirshberg, 2011).

Further, research has demonstrated that conversational partners match the pitch of their backchannels to the last segment of their partners' speech more than they match the pitch of their non-backchannels (Heldner, Edlund, Hirschberg, 2010). In the current study, I excluded all backchannels from analysis due to the constraint of the single-channel audio recordings. However, given the suggested significance of the prosodic entrainment of backchannels, it is necessary that future research measure this phenomenon for a potentially more accurate measurement of prosodic speech entrainment.

In addition, a further analysis of prosodic speech entrainment between romantic partners should include measurements of entrainment on a session basis. The current study calculates degree of entrainment exclusively on a turn-by-turn basis, measuring entrainment by comparing the prosodic similarities of adjacent IPUs at turn exchanges. However, it is also possible to measure acoustic-prosodic entrainment by looking at the overall matching of speakers' prosody throughout a conversation—known as entrainment on a session basis (Levitan & Hirshberg, 2011). While both of these methods are reliable measures of speech entrainment, conversational partners may exhibit speech entrainment at one level and not the other; for example, speakers who do not match their partners' prosody at turn exchanges (and therefore do not demonstrate turn-by-turn basis entrainment) may still display significant prosodic speech entrainment on a session basis due to exhibiting prosodic features with a mean center similar to that of their partner (Levitan & Hirshberg, 2011). Given that measuring entrainment on a turn-by-turn

basis only takes into account a very small portion of speech from a conversation (adjacent IPUs at turn exchanges), it is crucial that future studies on speech entrainment between romantic partners also measure entrainment on a session basis. This way, research will be able to provide a more comprehensive measurement of the extent to which romantic partners prosodically entrain to one another.

Furthermore, it is essential that studies investigating speech entrainment within romantic relationships examine the phenomenon through a combination of measurements of lexical, phonetic, and acoustic-prosodic matching. Since acoustic-prosodic entrainment takes place at the lowest, and most subconscious, level, it is possible that romantic partners demonstrate more variability in speech entrainment as measured by higher and more conceptual levels of lexical entrainment. Future work can assess lexical entrainment using LSM, which has already been shown to predict relationship stability (Ireland et al., 2011), in regards to the social factors examined in the current study (length or relationship, health of relationship, content of conversation). Moreover, by using a combination of these measurements of lexical, phonetic, and acoustic-prosodic entrainment, future studies could provide insight into the mechanisms behind entrainment between romantic partners. To exemplify, this method would allow researchers to investigate whether romantic partners entrain their speech significantly more at a higher conceptual or lower perceptual level, and whether this pattern is different from that of strangers.

Conclusion

The current study was designed to observe acoustic-prosodic speech entrainment in romantic relationships, and investigate the extent to which a couples' level of speech entrainment is dependent on factors such as the length of their relationship, the health of their relationship, and the content of their conversation. Results demonstrated that couples exhibited significant prosodic entrainment at both a local and global level. However, I found no significant relationships between the examined factors and the level of couples' speech entrainment. I suggest that these results provide evidence that acoustic-prosodic entrainment is more relevant to small-scale interactions between partners, rather that the larger social factors considered in the current study. Additionally, this study has many implications that may provide researchers with a more thorough framework for future work examining speech entrainment in romantic relationships.

APPENDIX A

Dyadic Functioning Scales

Dyadic Conflict Resolution:

This scale assesses the ability of the couple to work together to make decisions or resolve conflict in a manner that leads to mutual satisfaction. This scale is dyadic, and thus it should capture the essence of the interaction between partners. For instance, does the overall interaction style undermine one partner's self-confidence, is one partner dominant in making decisions or in conversation, or are both partners full participants in the interactive process?

This scale assesses the degree to which each partner's perspective is fully served by the process of resolving conflict, and whether the outcome is mutually satisfactory. Throughout the interaction, couple's conflict resolution ability should be scored at the manifest level. When discussing a topic that is a problem in their relationship, the assessment of a couple's ability to resolve a conflict should reflect observed interaction regardless of the issue a couple chooses to discuss.

In examining whether this couple is working together in a way that is mutually satisfying, there are several elements of the session to consider: when asked to discuss a topic that is a problem in their relationship, are they able to discuss each side openly, listen to each other's perspective, and come to a resolution that is mutually satisfying? Is the couple able to work cooperatively or does one partner dominate? Clearly couples will utilize many different tactics and behaviors in the service of working together toward a given goal. Furthermore, some couples will resolve conflict with little effort, whereas for others, the process may be quite strenuous.

Within couples receiving high scores, both partners should appear satisfied with the process of decision-making, whether it requires a lot or a little effort to resolve disagreements. A low score may be given to couples in which one partner appears satisfied with the process of conflict resolution, while the other partner is passively or actively dissatisfied with the dynamic. For such couples, decisions may actually appear to require less effort than for a higher scoring pair.

1. Very low or no satisfaction with the decision-making process. Partners may not make any effort to work toward goals, or maybe the only agreed-upon goal is to finish the task quickly. They may seem uninvolved with each other, interacting only when absolutely necessary and giving the impression that the interaction is awkward and difficult to coordinate, requiring a great deal of effort. Alternatively, partners may be at odds in working toward the task goals such that the process is painful and there is little to no satisfaction with the final outcome.

2. Low satisfaction with the decision-making process. Perhaps the only agreed-

upon goal is to finish the task quickly. There may be little discussion or little consideration for each other's feelings or the task. Brief collaboration may be seen, but interactions are often awkward and there is little sensitivity to the other person's perspective. Competition may be a problem, in that outcomes are won rather than achieved through compromise.

3. Low to moderate satisfaction. Partners make an adequate effort at collaborating, and generally manage to work together to pursue the goals of the session. Still, there is a lack of sensitivity and there is low satisfaction in the process or the outcome of the tasks. There may be times when insensitivity or competitiveness interferes with task goals. Alternatively, a couple may earn a 3 if there is little data to go by because opinions are being withheld. Overall, this couple manages to cooperate effectively during a good portion of the session.

4. Moderate satisfaction. Partners make some effort at collaborating, and are able to work together to achieve the goals of the session. However, these couples may not reveal much conflict because one or both partners tend to avoid challenging the other when there is the potential for conflict. Alternatively, a couple may be given this score if both partners are verbal and discuss their differing opinions, but the process seems to be somewhat strained or tense and leads to a less than satisfactory outcome.

5. Good satisfaction. Partners make a good effort at working together, and the majority of outcomes are agreeable to both partners. The process of getting there may be slightly strained; for instance, one partner may put him or herself down a few times during the process of coming to agree with the other. But overall, there is evidence of good cooperation and mutual pleasure.

6. High satisfaction. Partners work well together, and both partners feel good about the process of resolving conflict and about the outcome. The process is mature and cooperative. There is a high level of understanding and sensitivity between partners. There may be slight evidence of holding back or one partner may occasionally appear slightly dominant.

7. Very High satisfaction. Couples who earn a score of 7 are able to elaborate their positions on topics that come up during tasks, and come to resolutions using a mature and cooperative process. There is a high level of understanding and sensitivity between partners. These couples are able to communicate openly, and clearly respect each other’s opinions.

Overall Quality:

This scale is meant to serve as an outcome variable and as a predictor of subsequent functioning. We are evaluating how good this relationship seems to be and we expect scores will predict later success in relationships to some degree. High scores imply that this is a good relationship. We would be pleased if our son, daughter, niece, brother, friend, etc., were to be involved in a relationship of this quality. A high-quality relationship is characterized by: mutual caring, trust, and emotional closeness; sensitivity to one another's needs and wishes; deep sharing of experience; enjoyment of each other; and some potential for mutual understanding. The relationship may not necessarily have the mutual understanding and commitment that can only evolve over years of relating and, for contextual reasons, it may even be questionable that the relationship can survive. Nonetheless, given these individuals' respective developmental periods (mostly twenties and thirties), and given the circumstances of their lives, the relationship is as good in qualitative terms as could be expected. It is viewed as clearly growth enhancing ("enabling" in Hauser's terms), helping to prepare the individuals to move toward fully committed and positive relationships in the subsequent period. We would predict, unequivocally, that the participants will be able to have a fully committed, deep adult relationship.

Couples scores at the low end of this scale would either be devoid of the aforementioned qualities or are characterized by strikingly negative features (victimization, chronic intense conflict, rigidity of roles). The relationship could either be bland and empty or hurtful to one party or the other. Moderate scores can be various combinations of insufficiently present positive qualities or negative qualities present to some degree.

1. This is a very poor relationship. Either so little is present that one might question the salience of the relationship or it is a clearly exploitative or destructive relationship.

2. This is not a growth-enhancing relationship. Some positive features are present, but the relationship certainly is not very evolved. It may serve the individuals in a very limited way, or else negative features clearly outweigh the positive features.

3. One questions whether, on balance, this relationship is growth-enhancing. Either it just doesn't have enough positive features, or the negative features somewhat outweigh the positive features.

4. This relationship is still more positive than negative, and on the whole, it is a growth-enhancing relationship. More careful weighing of pros and cons is required to reach this judgment, however, than is true for a score of 5. Certain qualities may be lacking. There may be some limitation regarding sharing or experience or it may seem somewhat bland. If negative features are present (e.g.,

asymmetry or some inflexibility in roles), these remain overbalanced by positive qualities.

5. One's overall impression remains that this is a good relationship. Compared to higher scores, positive features are somewhat diminished, or occasional negative features are more clear. There is still no question about positive features clearly outweighing the negative.

6. This is a very good relationship. It is only that some isolated positive feature is insufficiently present, or some isolated concern arises as one observes the couple interact. Some lack of depth may be noted (perhaps because the relationship is new), or there may be some hint of a constraining or shared experience. Alternatively, there may be a hint of asymmetry or some imperfection in the process of conflict resolution.

7. Within the limits of capabilities of people in their twenties (or thirties), this is a wonderful relationship. The relationship is characterized by deep caring and love. There is no question regarding the emotional investment present. All of the qualities of a positive relationship are obvious. No concerns about the relationship arise.

APPENDIX B

Interview Collaboration Scale

This scale assesses how much couples collaborate when telling their relationship history together, how much stories are benefited by being told by both partners, and whether partners show a joint orientation in describing the relationship. High scores are given to couples that consistently work together to talk about the relationship, and build off each other to tell more complex stories. Individuals in high scoring couples clearly view the relationship in the same way, and relationship stories are stronger and richer for being told by both people. Low scores are given to couples that do not work together at all to tell relationship stories and talk about the relationship. Individuals in low scoring couples may often invalidate and disagree with what the other partner says.

1.Individuals within the couple do not work together at all to tell relationship stories, either telling similar stories in parallel without building off of what the partner says, or disagreeing and telling very different stories.

2.Partners occasionally collaborate in their storytelling, but typically tell stories alone without using the other partner as a tool to add depth to the storytelling.

3.The couple displays a mix of collaborative engagement and disagreement in the storytelling process. They both work together and use each other to tell more complex stories, but also sometimes invalidate each other or tell separate and unrelated stories.

4.The couple works well together to tell stories about their relationship, and are almost always collaborative. However there are occasional moments when partners invalidate each other or tell stories as individuals instead of working together.

5.The couple consistently works together to talk about the relationship, and the partners build off of what the other says to tell very rich and complex stories. These individuals clearly view the relationship in the same way, and relationship stories are stronger and richer for being told by both partners.

REFERENCES

Acitelli, L. K., Kenny, D. A., & Weiner, D. (2001). The importance of similarity and understanding of partners' marital ideals to relationship satisfaction. *Personal Relationships*, *8*(2), 167-185.

Anderson, C., Keltner, D., & John, O. P. (2003). Emotional convergence between people over time. *Journal of Personality and Social Psychology*, *84*(5), 1054.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189.

Bargh, J. A., Chen, M., & Burrows, M. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*(2), 230-244.

Boersma, P., & Weenink, D. (2012). Praat, version 5.5.

Bock, E. R. (2012). *Common Ground: A Look at Entrainment in Romantic Relationships* (Doctoral dissertation). Retrieved from The New School for Social Research of the New School.

Chartrand, T. L., & Bargh, J. A. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology*, *71*(3), 464-478.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893-910.

Cox, M. J. (1991). Marital and parent-child relationships study. Unpublished manuscript, University of North Carolina, Chapel Hill, NC.

Giles, H., Coupland, J., & Coupland, N. (Eds.). (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. New York: Cambridge University Press.

Haydon, K.C., Jonestrask, C., Guhn-Knight, H., & Salvatore, J.E. (under review). Adult attachment and conflict behavior predict distinct patterns of romantic conflict recovery.

Haydon, K. C., Roisman, G. I., & Burt, K. B. (2012). In search of security: The latent structure of the Adult Attachment Interview revisited. *Development and Psychopathology*, *24*(02), 589-606.

Heldner, M., Edlund, J., & Hirschberg, J. B. (2010). Pitch similarity in the vicinity of backchannels. In *Proceedings of Interspeech 2010,* Makuhari, Japan.

Honeycutt, J. M., & Cantrill, J. G. (2014). *Cognition, communication, and romantic relationships*. New York: Routledge

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39-44.

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

Lee, C. C., Black, M., Katsamanis, A., Lammert, A. C., Baucom, B. R., Christensen, A., Georgiou, P. G., & Narayanan, S. S. (2010). Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In *Proceedings of Interspeech 2010,* Makuhari, Japan.

Lee, C. C., Katsamanis, A., Black, M. P., Baucom, B. R., Christensen, A., Georgiou, P. G., & Narayanan, S. S. (2014). Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions. *Computer Speech & Language*, *28*(2), 518-539.

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech, 2011,* Florence, Italy.

Levitan, R., Gravano, A., & Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short papers*, Portland, OR.

Levitan, R., Gravano, A., Willson, L., Benus, S., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the*

Association for Computational Linguistics: Human Language Technologies, Atlanta, GA.

Lubold, N., & Pon-Barry, H. (2014). Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, Instanbul, Turkey.

Meeks, B. S., Hendrick, S. S., & Hendrick, C. (1998). Communication, love and relationship satisfaction. *Journal of Social and Personal Relationships*, *15*(6), 755-773.

Manson, J.H., Bryant, G.A., Gervais, M.M., & Kline, M. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior 34*(6), 419-426.

Nenkova, A., Gravano, A., Hirschberg, J. (2008). High Frequency Word Entrainment in Spoken Dialogue. In *Proceedings of ACL-08: Short Papers*, Columbus OH.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 169-190.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Price, R. A., & Vandenberg, S. G. (1980). Spouse similarity in American and Swedish couples. *Behavior genetics*, *10*(1), 59-71.

Reitter, D., & Moore, J. D. (2007). Predicting Success in Dialogue. In *Proceedings of 45th Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, *1*(4), 285-293.