Acoustic-Prosodic Entrainment and Social Behavior

Rivka Levitan[1], **Agustín Gravano**[2], **Laura Willson**[1],
Štefan Beňuš[3], **Julia Hirschberg**[1], **Ani Nenkova**[4]

[1] Dept. of Computer Science, Columbia University, New York, NY 10027, USA
[2] Departamento de Computación (FCEyN), Universidad de Buenos Aires, Argentina
[3] Constantine the Philosopher University & Institute of Informatics, Slovak Academy of Sciences, Slovakia
[4] Dept. of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104, USA

rlevitan@cs.columbia.edu, gravano@dc.uba.ar, law2142@barnard.edu,
sbenus@ukf.sk, julia@cs.columbia.edu, nenkova@seas.upenn.edu

Abstract

In conversation, speakers have been shown to entrain, or become more similar to each other, in various ways. We measure entrainment on eight acoustic features extracted from the speech of subjects playing a cooperative computer game and associate the degree of entrainment with a number of manually-labeled social variables acquired using Amazon Mechanical Turk, as well as objective measures of dialogue success. We find that male-female pairs entrain on all features, while male-male pairs entrain only on particular acoustic features (intensity mean, intensity maximum and syllables per second). We further determine that entrainment is more important to the perception of female-male social behavior than it is for same-gender pairs, and it is more important to the smoothness and flow of male-male dialogue than it is for female-female or mixed-gender pairs. Finally, we find that entrainment is more pronounced when intensity or speaking rate is especially high or low.

1 Introduction

Entrainment, also termed *alignment*, *adaptation*, *priming* or *coordination*, is the phenomenon of conversational partners becoming more similar to each other in what they say, how they say it, and other behavioral phenomena. Entrainment has been shown to occur for numerous aspects of spoken language, including speakers' choice of referring expressions (Brennan & Clark, 1996); linguistic style (Niederhoffer & Pennebaker, 2002; Danescu-Niculescu-Mizil et al., 2011); syntactic structure (Reitter et al., 2006); speaking rate (Levitan & Hirschberg, 2011); acoustic/prosodic features such as fundamental frequency, intensity, voice quality (Levitan & Hirschberg, 2011); and phonetics (Pardo, 2006).

Entrainment in many of these dimensions has also been associated with different measures of dialogue success. For example, Chartrand and Bargh (1999) demonstrated that mimicry of posture and behavior led to increased liking between the dialogue participants as well as a smoother interaction. They also found that naturally empathetic individuals exhibited a greater degree of mimicry than did others. Nenkova et al. (2008) found that entrainment on high-frequency words was correlated with naturalness, task success, and coordinated turn-taking behavior. Natale (1975) showed that an individual's social desirability, or "propensity to act in a social manner," can predict the degree to which that individual will match her partner's vocal intensity. Levitan et al. (2011) showed that entrainment on backchannel-preceding cues is correlated with shorter latency between turns, fewer interruptions, and a higher degree of task success. In a study of married couples discussing problems in their relationships, Lee et al. (2010) found that entrainment measures derived from pitch features were significantly higher in positive interactions than in negative interactions and were predictive of the polarity of the participants' attitudes.

These studies have been motivated by theoretical models such as Giles' Communication Accommodation Theory (Giles & Coupland, 1991), which proposes that speakers promote social approval or

2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 11–19,
Montréal, Canada, June 3-8, 2012.

efficient communication by adapting to their interlocutors' communicative behavior. Another theory informing the association of entrainment and dialogue success is the coordination-rapport hypothesis (Tickle-Degnen & Rosenthal, 1990), which posits that the degree of liking between conversational partners should be correlated with the degree of nonverbal coordination between them.

Motivated by such theoretical proposals and empirical findings, we hypothesized that entrainment on acoustic/prosodic dimensions such as pitch, intensity, voice quality and speaking rate might also be correlated with positive aspects of perceived social behaviors as well as other perceived characteristics of efficient, well-coordinated conversations. In this paper we describe a series of experiments investigating the relationship between objective acoustic/prosodic dimensions of entrainment and manually-annotated perception of a set of social variables designed to capture important aspects of conversational partners' social behaviors. Since prior research on other dimensions of entrainment has sometimes observed differences in degree of entrainment between female-female, male-male and mixed gender groups (Bilous & Krauss, 1988; Pardo, 2006; Namy et al., 2002), we also examined our data for variation by gender pair, considering female-female, male-male, and female-male pairs of speakers separately. If previous findings extend to acoustic/prosodic entrainment, we would expect female-female pairs to entrain to a greater degree than male-male pairs and female partners in mixed gender pairs to entrain more than their male counterparts. Since prior findings posit that entrainment leads to smoother and more natural conversations, we would also expect degree of entrainment to correlate with perception of other characteristics descriptive of such conversations.

Below we describe the corpus and annotations used in this study and how our social annotations were obtained in Sections 2 and 3. We next discuss our method and results for the prevalence of entrainment among different gender groups (Section 4). In Sections 5 and 6, we present the results of correlating acoustic entrainment with social variables and objective success measures, respectively. Finally, in Section 7, we explore entrainment in cases of outlier feature values.

2 The Columbia Games Corpus

The Columbia Games Corpus (Gravano & Hirschberg, 2011) consists of approximately nine hours of spontaneous dialogue between pairs of subjects playing a series of computer games. Six females and seven males participated in the collection of the corpus; eleven of the subjects returned on a different day for another session with a new partner.

During the course of each session, a pair of speakers played three Cards games and one Objects game. The work described here was carried out on the Objects games. This section of each session took 7m 12s on average. We have a total of 4h 19m of Objects game speech in the corpus.

For each task in an Objects game, the players saw identical collections of objects on their screens. However, one player (the Describer) had an additional target object positioned among the other objects, while the other (the Follower) had the same object at the bottom of her screen. The Describer was instructed to describe the position of the target object so that the Follower could place it in exactly the same location on her screen. Points (up to 100) were awarded based on how well the Follower's target location matched the describers. Each pair of partners completed 14 such tasks, alternating roles with each task. The partners were separated by a curtain to ensure that all communication was oral.

The entire corpus has been orthographically transcribed and words aligned with the speech source. It has also been ToBI-labeled (Silverman et al., 1992) for prosodic events, as well as labeled for turn-taking behaviors.

3 Annotation of Social Variables

In order to study how entrainment in various dimensions correlated with perceived social behaviors of our subjects, we asked Amazon Mechanical Turk[1] annotators to label the 168 Objects games in our corpus for an array of social behaviors perceived for each of the speakers, which we term here "social variables."

Each Human Intelligence Task (HIT) presented to the AMT workers for annotation consisted of a single Objects game task. To be eligible for our HITs,

[1] http://www.mturk.com

annotators had to have a 95% success rate on previous AMT HITs and to be located in the United States. They also had to complete a survey establishing that they were native English speakers with no hearing impairments. The annotators were paid $0.30 for each HIT they completed. Over half of the annotators completed fewer than five hits, and only four completed more than twenty.

The annotators listened to an audio clip of the task, which was accompanied by an animation that displayed a blue square or a green circle depending on which speaker was currently talking. They were then asked to answer a series of questions about each speaker: *Does Person A/B believe s/he is better than his/her partner? Make it difficult for his/her partner to speak? Seem engaged in the game? Seem to dislike his/her partner? Is s/he bored with the game? Directing the conversation? Frustrated with his/her partner? Encouraging his/her partner? Making him/herself clear? Planning what s/he is going to say? Polite? Trying to be liked? Trying to dominate the conversation?* They were also asked questions about the dialogue as a whole: *Does it flow naturally? Are the participants having trouble understanding each other? Which person do you like more? Who would you rather have as a partner?*

A series of check questions with objectively determinable answers (e.g. "Which speaker is the Describer?") were included among the target questions to ensure that the annotators were completing the task with integrity. HITs for which the annotator failed to answer the check questions correctly were disqualified.

Each task was rated by five unique annotators who answered "yes" or "no" to each question, yielding a score ranging from 0 to 5 for each social variable, representing the number of annotators who answered "yes." A fuller description of the annotation for social variables can be found in (Gravano et al., 2011).

In this study, we focus our analysis on annotations of four social variables:

- Is the speaker trying to be liked?
- Is the speaker trying to dominate the conversation?
- Is the speaker giving encouragement to his/her partner?
- Is the conversation awkward?

We correlated annotations of these variables with an array of acoustic/prosodic features.

4 Acoustic entrainment

We examined entrainment in this study in eight acoustic/prosodic features:

- Intensity mean
- Intensity max
- Pitch mean
- Pitch max
- Jitter
- Shimmer
- Noise-to-harmonics ratio (NHR)
- Syllables per second

Intensity is an acoustic measure correlated with perceived loudness. Jitter, shimmer, and noise-to-harmonics ratios are three measures of voice quality. Jitter describes varying pitch in the voice, which is perceived as a rough sound. Shimmer describes fluctuation of loudness in the voice. Noise-to-harmonics ratio is associated with perceived hoarseness. All features were speaker-normalized using *z*-scores.

For each task, we define entrainment between partners on each feature *f* as

$$ENT_p = -|\text{speaker1}_f - \text{speaker2}_f|$$

where speaker[1,2]_f represents the corresponding speaker's mean for that feature over the task.

We say that the corpus shows evidence of entrainment on feature *f* if ENT_p, the similarities between partners, are significantly greater than ENT_x, the similarities between non-partners:

$$ENT_x = -\frac{\sum_i |\text{speaker1}_f - \text{X}_{i,f}|}{|\text{X}|}$$

where X is the set of speakers of same gender and role as the speaker's partner who are not paired with the speaker in any session. We restrict the comparisons to speakers of the same gender and role as the speaker's partner to control for the fact that differences may simply be due to differences in gender or role. The results of a series of paired *t*-tests comparing ENT_p and ENT_x for each feature are summarized in Table 1.

Feature	FF	MM	FM
Intensity mean	✓	✓	✓
Intensity max	✓	✓	✓
Pitch mean			✓
Pitch max			✓
Jitter	✓		✓
Shimmer	✓		✓
NHR			✓
Syllables per sec	✓	✓	✓

Table 1: Evidence of entrainment for gender pairs. A tick indicates that the data shows evidence of entrainment on that row's feature for that column's gender pair.

We find that **female-female** pairs in our corpus entrain on, in descending order of significance, jitter, intensity max, intensity mean, syllables per second and shimmer. They do not entrain on pitch mean or max or NHR. **Male-male** pairs show the least evidence of entrainment, entraining only on intensity mean, intensity max, and syllables per second, supporting the hypothesis that entrainment is less prevalent among males. **Female-male** pairs entrain on, again in descending order of significance, intensity mean, intensity max, jitter, syllables per second, pitch mean, NHR, shimmer, and pitch max – in fact, on every feature we examine, with significance values in each case of p<0.01.

To look more closely at the entrainment behavior of males and females in mixed-gender pairs, we define $ENT2_p$ as follows:

$$ENT2_p = -\frac{\sum_i |\mathrm{P}_{i,f} - \mathrm{T}_{i,f}|}{|\mathrm{T}|}$$

where T is the set of the pause-free chunks of speech that begin a speaker's turns, and P is the corresponding set of pause-free chunks that end the interlocutor's preceding turns. Unlike ENT_p, this measure is asymmetric, allowing us to consider each member of a pair separately.

We compare $ENT2_p$ for each feature for males and females of mixed gender pairs. Contrary to our hypothesis that females in mixed-gender pairs would entrain more, we found no significant differences in partner gender. Females in mixed-gender pairs do not match their interlocutor's previous turn any more than do males. This may be due to the fact that, as shown in Table 1, the overall differences between partners in mixed-gender pairs are quite low, and so neither partner may be doing much turn-by-turn matching.

Feature	FM	MM	F	*p*
Intensity mean	↑	↓	3.83	0.02
Intensity max	↑	↓	4.01	0.02
Syllables per sec	↓	↓	2.56	0.08

Table 2: Effects of gender pair on entrainment. An arrow pointing up indicates that the group's normalized entrainment for that feature is greater than that of female-female pairs; an arrow pointing down indicates that it is smaller.

However, as we expected, entrainment is least prevalent among male-male pairs. Although we expected female-female pairs to exhibit the highest prevalence of entrainment, they do not show evidence of entrainment on pitch mean, pitch max or NHR, while female-male pairs entrain on every feature. In fact, although ENT_p for these features is not significantly smaller between female-female pairs than between female-male pairs, ENT_x, the overall similarity among non-partners for these features, is significantly larger between females than between females and males. The degree of similarity between female-female partners is therefore attributable to the overall similarity between females rather than the effect of entrainment.

All three types of pairs exhibit entrainment on intensity mean, intensity max, and syllables per second. We look more closely into the gender-based differences in entrainment behavior with an ANOVA with the ratio of ENT_p to ENT_x as the dependent variable and gender pair as the independent variable. Normalizing ENT_p by ENT_x allows us to compare the degree of entrainment across gender pairs. Results are shown in Table 2. Male-male pairs have lower entrainment than female-female pairs for every feature; female-male pairs have higher entrainment than female-female pairs for intensity mean and max and lower for syllables per second ($p < 0.1$). These results are consistent with the general finding that male-male pairs entrain the least and female-male pairs entrain the most.

5 Entrainment and social behavior

We next correlate each of the social variables described in Section 3 with ENT_p for our eight acoustic features. Based on Communication Accommodation Theory, we would expect *gives encouragement*, a variable representing a desirable social characteristic, to be positively correlated with entrainment. Conversely, *conversation awkward* should be negatively correlated with entrainment. We note that *Trying to be liked* is negatively correlated with the *like more* variable in our data – that is, annotators were less likely to prefer speakers whom they perceived as trying to be liked. This reflects the intuition that someone overly eager to be liked may be perceived as annoying and socially inept. However, similarity-attraction theory states that similarity promotes attraction, and someone might therefore entrain in order to obtain his partner's social approval. This idea is supported by Natale's finding that the need for social approval is predictive of the degree of a speaker's convergence on intensity (Natale, 1975). We can therefore expect *trying to be liked* to positively correlate with entrainment. Speakers who are perceived as *trying to dominate* may be *overly* entraining to their interlocutors in what is sometimes called "dependency overaccommodation." Dependency overaccommodation causes the interlocutor to appear dependent on the speaker and gives the impression that the speaker is controlling the conversation (West & Turner, 2009).

The results of our correlations of social variables with acoustic/prosodic entrainment are generally consonant with these intuitions. Although it is not straightforward to compare correlation coefficients of groups for which we have varying amounts of data, for purposes of assessing trends, we will consider a correlation strong if it is significant at the $p < 0.00001$ level, moderate at the $p < 0.01$ level, and weak at the $p < 0.05$ level. The results are summarized in Table 3; we present only the significant results for space considerations.

For **female-female pairs**, *giving encouragement* is weakly correlated with entrainment on intensity max and shimmer. *Conversation awkward* is weakly correlated with entrainment on jitter. For **male-male pairs**, *trying to be liked* is moderately correlated with entrainment on intensity mean and weakly correlated with entrainment on jitter and NHR. *Giving encouragement* is moderately correlated with entrainment on intensity mean, intensity max, and NHR. For **female-male pairs**, *trying to be liked* is moderately correlated with entrainment on pitch mean. *Giving encouragement* is strongly correlated with entrainment on intensity mean and max and moderately correlated with entrainment on pitch mean and shimmer. However, it is *negatively* correlated with entrainment on jitter, although the correlation is weak. *Conversation awkward* is weakly correlated with entrainment on jitter.

As we expected, *giving encouragement* is correlated with entrainment for all three gender groups, and *trying to be liked* is correlated with entrainment for male-male and female-male groups. However, *trying to dominate* is not correlated with entrainment on any feature, and *conversation awkward* is actually positively correlated with entrainment on jitter.

Entrainment on jitter is a clear outlier here, with all of its correlations contrary to our hypotheses. In addition to being positively correlated with *conversation awkward*, it is the only feature to be negatively correlated with *giving encouragement*.

Entrainment is correlated with the most social variables for female-male pairs; these correlations are also the strongest. We therefore conclude that acoustic entrainment is not only most prevalent for mixed-gender pairs, it is also more important to the perception of female-male social behavior than it is for same-gender pairs.

6 Entrainment and objective measures of dialogue success

We now examine acoustic/prosodic entrainment in our corpus according to four objective measures of dialogue success: the mean latency between turns, the percentage of turns that are interruptions, the percentage of turns that are overlaps, and the number of turns in a task.

High latency between turns can be considered a sign of an unsuccessful conversation, with poor turn-taking behavior indicating a possible lack of rapport and difficulty in communication between the partners. A high percentage of interruptions, another example of poor turn-taking behavior, may be a symptom of or a reason for hostility or awkwardness be-

Social	Acoustic	df	*r*	*p*
Female-Female				
Giving enc.	Int. max		-0.24	0.03
	Shimmer		-0.24	0.03
Conv. awkward	Jitter		-0.23	0.03
Male-Male				
Trying to be liked	Int. mean		-0.30	0.006
	Jitter		-0.27	0.01
	NHR		-0.23	0.03
Giving enc.	Int. mean		-0.39	0.0003
	Int. max		-0.31	0.005
	NHR		-0.30	0.005
Female-Male				
Trying to be liked	Pitch mean		-0.26	0.001
Giving enc.	Int. mean		-0.36	2.8e-06
	Int. max		-0.31	7.7e-05
	Pitch mean		-0.23	0.003
	Jitter		0.19	0.02
	Shimmer		-0.16	0.04
Conv. awkward	Jitter		-0.17	0.04

Table 3: Correlations between entrainment and social variables.

tween partners. We expect these measures to be negatively correlated with entrainment. Conversely, a high percentage of overlaps may be a symptom of a well-coordinated conversation that is flowing easily. In the guidelines for the turn-taking annotation of the Games Corpus (Gravano, 2009), overlaps are defined as cases in which Speaker 2 takes the floor, overlapping with the completion of Speaker 1's utterance. Overlaps require the successful reading of turn-taking cues and by definition preclude awkward pauses. We expect a high percentage of overlaps to correlate positively with entrainment.

The number of turns in a task can be interpreted either positively or negatively. A high number is negative in that it is the sign of an inefficient dialogue, one which takes many turn exchanges to accomplish the objective. However, it may also be the sign of easy, flowing dialogue between the partners. In our domain, it may also be a sign of a high-achieving pair who are placing the object meticulously in order to secure every single point. We therefore expect the number of turns to be positively correlated with entrainment. As before, we consider a correlation strong if it is significant at the $p < 0.00001$ level, moderate at the $p < 0.01$ level, and weak at the $p < 0.05$ level. The significant correlations are presented in Table 4.

Objective	Acoustic	df	*r*	*p*
Female-Female				
Latency	Int. mean		0.22	0.04
	Int. max		0.31	0.005
	Pitch mean		0.24	0.02
	Jitter		0.29	0.007
	Shimmer		0.33	0.002
	Syllables/sec		0.39	0.0002
# Turns	Int. max		-0.30	0.006
	Shimmer		-0.34	0.002
	NHR		-0.24	0.03
	Syllables/sec		-0.28	0.01
% Overlaps	Int. max		-0.23	0.04
	Shimmer		-0.30	0.005
% Interruptions	Shimmer		-0.33	0.005
Male-Male				
Latency	Int. mean		0.57	8.8e-08
	Int. max		0.43	0.0001
	Pitch mean		0.52	2.4e-06
	Pitch max		0.61	5.7e-09
	Jitter		0.65	4.5e-10
	NHR		0.40	0.0004
# Turns	Int. mean		-0.29	0.0002
	Pitch mean		-0.32	0.003
	Pitch max		-0.29	0.007
	NHR		-0.47	7.9e-06
	Syllables/sec		-0.25	0.02
% Overlaps	Int. mean		-0.39	0.0002
	Int. max		-0.39	0.0002
% Interruptions	NHR		-0.33	0.002
Female-Male				
# Turns	Int. mean		-0.24	0.003
	Int. max		-0.19	0.02
	Shimmer		-0.16	0.04
% Overlaps	Shimmer		-0.26	0.001

Table 4: Correlations between entrainment and objective variables.

For **female-female pairs**, mean latency between turns is *negatively* correlated with entrainment on all variables except pitch max and NHR. The correlations are weak for intensity mean and pitch mean and moderate for intensity max, jitter, shimmer, and syllables per second. The number of turns is moderately correlated with entrainment on intensity max and shimmer and weakly correlated with entrainment on syllables per second. Contrary to our expectations, the percentage of interruptions is *positively* (though moderately) correlated with entrainment on shimmer; the percentage of overlaps is moderately correlated with entrainment on shimmer and weakly correlated with entrainment on intensity max.

Male-male pairs show the most correlations between entrainment and objective measures of dialogue success. The latency between turns is *negatively* correlated with entrainment on all variables except shimmer and syllables per second; the correlations are moderate for intensity max and NHR and strong for the rest. The number of turns in a task is *positively* correlated with entrainment on every variable except intensity mean, jitter and shimmer: strongly for NHR; moderately for intensity mean, pitch mean, and pitch max; and weakly for syllables per second.. The percentage of overlaps is moderately correlated with entrainment on intensity mean and max. The percentage of interruptions is moderately correlated with entrainment on NHR.

For **female-male pairs**, the number of turns is moderately correlated with entrainment on intensity mean and weakly correlated with entrainment on intensity max and shimmer. The percentage of overlaps is moderately correlated with entrainment on shimmer.

For the most part, the directions of the correlations we have found are in accordance with our hypotheses. Latency is negatively correlated with entrainment and overlaps and the number of turns are positively correlated. A puzzling exception is the percentage of interruptions, which is positively correlated with entrainment on shimmer (for female-female pairs) and NHR (for male-male pairs).

While the strongest correlations were for mixed-gender pairs for the social variables, we find that the strongest correlations for objective variables are for male-male pairs, which also have the greatest number of correlations. It therefore seems that while entrainment is more important to the perception of social behavior for mixed-gender pairs than it is for same-gender pairs, it is more important to the smoothness and flow of dialogue for male-male pairs than it is for female-female or female-male pairs.

7 Entrainment in outliers

Since acoustic entrainment is generally considered an unconscious phenomenon, it is interesting to consider tasks in which a particular feature of a person's speech is particularly salient. This will occur when a feature differs significantly from the norm – for example, when a person's voice is unusually loud or soft. Chartrand and Bargh (1999) suggest that the psychological mechanism behind the entrainment is the perception-behavior link, the finding that the act of observing another's behavior increases the likelihood of the observer's engaging in that behavior. Based on this finding, we hypothesize that a partner pair containing one "outlier" speaker will exhibit more entrainment on the salient feature, since that feature is more likely to be observed and therefore imitated.

We consider values in the 10th or 90th percentile for a feature "outliers." We can consider ENT_x, the similarity between a speaker and the speakers of her partner's role and gender with whom she is never paired, the "baseline" value for the similarity between a speaker and her interlocutor when no entrainment occurs. $ENT_p - ENT_x$, the difference between the similarity existing between partners and the baseline similarity, is then a measure of how much entrainment exists relative to baseline.

We compare $ENT_p - ENT_x$ for "normal" versus "outlier" speakers. ENT_p should be smaller for outlier speakers, since their interlocutors are not likely to be similarly unusual. However, ENT_x should also be lower for outlier speakers, since by definition they diverge from the norm, while the normal speakers by definition represent the norm. It is therefore reasonable to expect $ENT_p - ENT_x$ to be the same for outlier speakers and normal speakers.

If $ENT_p - ENT_x$ is *higher* for outlier speakers, that means that ENT_p is higher than we expect, and entrainment is *greater* relative to baseline for pairs containing an outlier speaker. If $ENT_p - ENT_x$ is *lower* for outlier speakers, that means that ENT_p is

Acoustic	*t*	df	*p*
Intensity mean	5.66	94.26	1.7e-07
Intensity max	8.29	152.05	5.5e-14
Pitch mean	-1.20	76.82	N.S.
Pitch max	-0.84	76.76	N.S.
Jitter	0.36	70.23	N.S.
Shimmer	2.64	102.23	0.02
NHR	-0.92	137.34	N.S.
Syllables per sec	2.41	72.60	0.02

Table 5: *T*-tests for relative entrainment for outlier vs. normal speakers.

lower than we expect, and pairs containing an outlier speaker entrain *less* than do pairs of normal speakers, even allowing for the fact that their usual values should be further apart to begin with.

The results for *t*-tests comparing $ENT_p - ENT_x$ for "normal" versus "outlier" speakers are shown in Table 5. Outlier pairs have *higher* relative entrainment than do normal pairs for intensity mean and max, shimmer, and syllables per second. This means that speakers confronted with an interlocutor who diverges widely from the norm for those four features make a *larger* adjustment to their speech in order to converge to that interlocutor.

An ANOVA shows that relative entrainment on intensity max is higher in outlier cases for male-male pairs than for female-female pairs and even higher for female-male pairs (F=11.33, p=5.3e-05). Relative entrainment on NHR in these cases is *lower* for male-male pairs than for female-female pairs and higher for female-male pairs (F=11.41, p=6.5e-05). Relative entrainment on syllables per second is lower for male-male pairs and higher for female-male pairs (F=5.73, p=0.005). These results differ slightly from the results in Table 2 for differences in entrainment in the general case among gender pairs, reinforcing the idea that cases in which feature values diverge widely from the norm are unique in terms of entrainment behavior.

8 Conclusion

Our study of entrainment on acoustic/prosodic variables yields new findings about entrainment behavior for female-female, male-male, and mixed-gender dyads, as well as the association of entrainment with perceived social characteristics and objective measures of dialogue smoothness and efficiency. We find that entrainment is the most prevalent for mixed-gender pairs, followed by female-female pairs, with male-male pairs entraining the least. Entrainment is the most important to the perception of social behavior of mixed-gender pairs, and it is the most important to the efficiency and flow of male-male dialogues.

For the most part, the directions of the correlations of entrainment with success variables accord with hypotheses motivated by the relevant literature. Giving encouragement and trying to be liked are positively correlated with entrainment, as are percentage of overlaps and number of turns. Mean latency, a symptom of a poorly-run conversation, is negatively associated with entrainment. However, several exceptions suggest that the associations are not straightforward and further research must be done to fully understand the relationship between entrainment, social characteristics and dialogue success. In particular, the explanation behind the associations of entrainment on certain variables with certain social and objective measures is an interesting direction for future work.

Finally, we find that in "outlier" cases where a particular speaker diverges widely from the norm for intensity mean, intensity max, or syllables per second, entrainment is more pronounced. This supports the theory that the perception-behavior link is the mechanism behind entrainment and provides a possible direction for research into why speakers entrain on certain features and not others. In future work we will explore this direction and go more thoroughly into individual differences in entrainment behavior.

Acknowledgments

This material is based upon work supported in part by NSF IIS-0307905, NSF IIS-0803148, UBACYT 20020090300087, ANPCYT PICT-2009-0026, CONICET, VEGA No. 2/0202/11; and the EUSF (ITMS 26240220060).

References

Amazon Mechanical Turk, *http://www.mturk.com*.

Frances R. Bilous and Robert M. Krauss 1988. Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication*, 8(3/4):183–194.

Susan E. Brennan and Herbert H. Clark. 1996. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22(6):1482–1493.

Tanya L. Chartrand and John A. Bargh. 1999. The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology*, 76(6):893–910.

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark My Words! Linguistic Style Accommodation in Social Media. *Proceedings of WWW 2011*.

H. Giles and N. Coupland. 1991. *Language: Contexts and Consequences.* Pacific Grove, CA: Brooks/Cole.

Agustín Gravano. 2009. *Turn-Taking and Affirmative Cue Words in Task-Oriented Dialogue.* Ph.D. thesis, Columbia University, New York.

Agustín Gravano and Julia Hirschberg. 2011. Turn-taking cues in task-oriented dialogue. *Computer Speech and Language*, 25(3):601–634.

Agustín Gravano, Rivka Levitan, Laura Willson, Štefan Beňuš, Julia Hirschberg, Ani Nenkova. 2011. Acoustic and Prosodic Correlates of Social Behavior. *Interspeech 2011*.

Chi-Chun Lee, Matthew Black, Athanasios Katsamanis, Adam Lammert, Brian Baucom, Andrew Christensen, Panayiotis G. Georgiou, Shrikanth Narayanan. 2010. Quantification of Prosodic Entrainment in Affective Spontaneous Spoken Interactions of Married Couples. *Eleventh Annual Conference of the International Speech Communication Association*.

Rivka Levitan, Agustín Gravano, and Julia Hirschberg. 2011. Entrainment in Speech Preceding Backchannels. *Proceedings of ACL/HLT 2011*.

Rivka Levitan and Julia Hirschberg. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech 2011*.

Laura L. Namy, Lynne C. Nygaard, Denise Sauerteig. 2002. Gender differences in vocal accommodation: the role of perception. *Journal of Language and Social Psychology*, 21(4):422–432.

Michael Natale. 1975. Convergence of Mean Vocal Intensity in Dyadic Communication as a Function of Social Desirability. *Journal of Personality and Social Psychology*, 32(5):790–804.

Ani Nenkova, Agustín Gravano, and Julia Hirschberg. 2008. High-frequency word entrainment in spoken dialogue. *Proceedings of ACL/HLT 2008*.

Kate G. Niederhoffer and James W. Pennebaker. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, 21(4):337–360.

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

David Reitter, Johanna D. Moore, and Frank Keller. 1996. Priming of Syntactic Rules in Task-Oriented Dialogue and Spontaneous Conversation. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*.

Kim Silverman, Mary Beckman, John Pitrelli, Mori Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, Julia Hirschberg. 1992. TOBI: A Standard for Labeling English Prosody. *ICSLP-1992*, 867-870.

Linda Tickle-Degnen and Robert Rosenthal. 1990. The Nature of Rapport and its Nonverbal Correlates. *Psychological Inquiry*, 1(4):285–293.

Richard West & Lynn Turner. 2009. *Introducing Communication Theory: Analysis and Application*. McGraw-Hill Humanities/Social Sciences/Languages, 4th edition.