Research Article

Rhythm Perception, Speaking Rate Entrainment, and Conversational Quality: A Mediated Model

Camille J. Wynn,[a] iD **Tyson S. Barrett,**[b] iD **and Stephanie A. Borrie**[a] iD

[a] Department of Communicative Disorders and Deaf Education, Utah State University, Logan, UT [b] Department of Psychology, Utah State University, Logan, UT

ABSTRACT

Purpose: Acoustic–prosodic entrainment, defined as the tendency for individuals to modify their speech behaviors to more closely align with the behaviors of their conversation partner, plays an important role in successful interaction. From a mechanistic perspective, acoustic–prosodic entrainment is, by its very nature, a rhythmic activity. Accordingly, it is highly plausible that an individual's rhythm perception abilities play a role in their ability to successfully entrain. Here, we examine the impact of rhythm perception in speaking rate entrainment and subsequent conversational quality.

Method: A round-robin paradigm was used to collect 90 dialogues from neurotypical adults. Additional assessments determined participants' rhythm perception abilities, social competence, and partner familiarity (i.e., whether the conversation partners knew each other prior to the interaction. Mediation analysis was used to examine the relationships between rhythm perception scores, speaking rate entrainment (using a measure of static local synchrony), and a measure of conversational success (i.e., conversational quality) based on third-party listener observations. Findings were compared to the same analysis with three additional predictive factors: participant gender, partner familiarity, and social competence.

Results: Results revealed a relationship between rhythm perception and speaking rate entrainment. In unfamiliar conversation partners, there was a relationship between speaking rate entrainment and conversational quality. The relationships between entrainment and each of the three additional factors (i.e., gender, partner familiarity, and social competence) were nonsignificant.

Conclusions: In unfamiliar conversation partners, better rhythm perception abilities were indicative of increased conversational quality mediated by higher levels of speaking rate entrainment. These results support theoretical postulations specifying rhythm perception abilities as a component of acoustic–prosodic entrainment, which, in turn, facilitates conversational success. Knowledge of this relationship contributes to the development of a causal framework for considering a mechanism by which rhythm perception deficits in clinical populations may impact conversational success.

Conversation involves much more than a static back-and-forth exchange of information. Rather, during conversation, partners engage in a rhythmic and dynamic interplay, continually altering their communicative behaviors

and coordinating these actions with one another. One well-established coordinative behavior is acoustic–prosodic entrainment, a phenomenon in which interlocutors modify their speech behaviors to align with the same behaviors of their partner. This alignment of behavior has been demonstrated in many different acoustic–prosodic elements of speech across articulatory (e.g., articulatory precision; Borrie, Wynn, et al., 2020; Lubold et al., 2019), phonatory (e.g., fundamental frequency; Borrie & Liss, 2014; Reichel et al., 2018), and prosodic (e.g., speech rate; Cohen-Priva et al., 2017; Wynn & Borrie, 2020b) aspects of speech.

Correspondence to Camille J. Wynn: cwynn214@gmail.com. **Disclosure:** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication. They also certify that they have no affiliations with or involvement in any organization or entity with any financial or nonfinancial interests in the subject matter or materials discussed in this article.*

According to the interactive alignment model, entrainment serves as an important grounding mechanism upon which conversation partners build shared mental representations, simplifying linguistic processing and reducing the computational load necessary for conversation (Pickering & Garrod, 2004). This, in turn, benefits the conversation. Research has indicated that entrainment is one strategy that people employ to increase the overall success of their interactions. For instance, entrainment is highly predictive of several functional measures of successful interaction including communicative efficiency (e.g., Borrie et al., 2015; Nenkova et al., 2008), cooperation (e.g., Manson et al., 2013; Taylor & Thomas, 2008), and collaboration (e.g., Polyanskaya et al., 2019a; Weidman et al., 2016). One important benefit of entrainment is its role in conversational quality or the overall *flow* of the interaction (e.g., Local, 2007; Wilson & Wilson, 2005). For example, Michalsky et al. (2018) found that higher levels of entrainment led to better perceptual ratings of conversational quality made by partners following the interaction. Presumably, greater conversational flow is associated with the perceived connection and rapport often reported by communication partners in highly entrained conversations (e.g., Aguilar et al., 2016; Schweitzer et al., 2017).

Although acoustic–prosodic entrainment has been demonstrated as a relatively pervasive communication phenomenon in the conversations of neurotypical individuals, there is a considerable amount of variability in the degree to which conversation partners entrain with one another and, subsequently, experience conversational success. Within the literature, sources of this variability—that is, the factors that may facilitate entrainment—have been widely discussed. One area of research has focused on factors that are theorized to impact an individual's *ability* to entrain. For example, Lewandowski and Jilka (2019) have argued that cognitive skills such as attention enhance or inhibit a person's ability to recognize, store, and retrieve the information regarding their partner's communicative behaviors. Beyond attention, the influence of working memory, social aptitude, and linguistic distance have been examined (Kim et al., 2011; Yu et al., 2013). A second area of research focuses on factors that are theorized to impact an individual's *motivation* to entrain. For example, a person's social desirability or rejection sensitivity may affect their drive to adjust their communication behaviors toward their partners to gain social acceptance and reduce social differences (Aguilar et al., 2016; Natale, 1975). Similarly, the conversational role of a speaker and their relative status (e.g., student vs. professional) within a given conversation may affect an individuals' desire for social approval and, thus, their motivation to entrain to the behaviors of their partner (Gregory & Webster, 1996; Pardo, 2006; Reichel et al., 2018; Street, 1984). Additional factors that may affect both *ability* and *motivation,* such as personality features, age, gender, partner familiarity, and relationship length, have also been explored (Cohen-Priva et al., 2017; Namy et al., 2002; Pardo et al., 2012; Weidman et al., 2016; Yu et al., 2013). Despite extensive research in this area, definitive conclusions are sparse. While some factors have been shown to predict entrainment, many have not. Additionally, there are contrasting results across different studies and contexts. For example, with regard to gender, some studies indicate higher levels of entrainment in male dyads (e.g., Pardo, 2006), whereas others have found higher rates in female dyads (e.g., Namy et al., 2002). Still, other studies have found no significant differences between male and female speakers (e.g., Yu et al., 2013). At present, there is no clear consensus regarding the factors that are most important for entrainment.

While the findings and ideas set forth by this body of research are valuable, there is one potential factor that has been surprisingly overlooked. From a theoretical perspective, entrainment of motoric behavior (including acoustic–prosodic entrainment) is, by its very nature, a rhythmic activity. Phillips-Silver et al. (2010), describe entrainment as "rhythmic responsiveness to a perceived rhythmic signal" (p. 3). Drawing upon the ideas of Todd et al. (2002), they note that entrainment therefore requires a robust sensory system by which rhythmic information can be extracted from the speech signal of a communication partner and subsequently integrated into their own speech patterns. Accordingly, it is highly plausible that an individual's rhythm perception abilities play a role in their capacity to entrain to their conversation partner. Indeed, rhythm perception abilities have been linked with greater success in other important aspects of communication including speech processing (e.g., Borrie et al., 2017; Slater & Krauss, 2016), stress perception, word segmentation (e.g., Hausen et al., 2013; Magne et al., 2016), and syntax use (e.g., Canette et al., 2019; Gordon et al., 2015). However, to date, the relationship between rhythm perception abilities and acoustic–prosodic entrainment in neurotypical populations has not been examined.

Despite the lack of an empirical link between rhythm perception and acoustic–prosodic entrainment in neurotypical populations, research from clinical populations with rhythm perception deficits provides evidence, at least indirectly, of such a relationship. For example, Lagrois et al. (2019) found that individuals with a beat processing disorder (i.e., deficits in tracking the beat in music) were unable to rhythmically tap their finger to natural spoken output. Presumably, their inability to perceive the rhythmic patterns of a given signal prohibited them from entraining their rhythmic output with the sensory input of the speech stimuli. Within the realm of verbal communication, lower levels of entrainment have been identified in the speech patterns of populations with rhythm perception deficits (Isenhower

Figure 1. Conceptual model employed to test the hypothesis that rhythmic ability effects conversational success through acoustic–prosodic entrainment. Additional factors including gender, partner familiarity, and social competence are also included within our model.
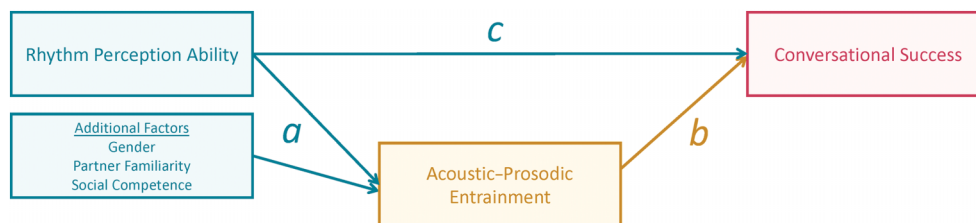
et al., 2012; Polonenko et al., 2017; Stabej et al., 2012; Tordjman et al., 2015) such as children with cochlear implants (Freeman & Pisoni, 2017) and autistic individuals (Lehnert-LeHouillier et al., 2020; Wynn et al., 2018). While neurotypical populations do not, as a general whole, exhibit rhythmic perception deficits to the extent exhibited within these clinical groups, a great deal of individual variability in rhythm perception ability does exist (e.g., Fujii & Schlaug, 2013; Grahn & Schuit, 2012; Law & Zentner, 2012). For example, in a cohort of 121 neurotypical individuals, Wallentin et al. (2010) found that scores on a validated rhythm perception task ranged from 100% accuracy to scores that do not rank significantly above chance. Given the large individual variability in rhythm perception coupled with strong theoretical reasoning and evidence from clinical populations, we advance the prediction that rhythm perception abilities may facilitate acoustic–prosodic entrainment in neurotypical individuals.

The purpose of this study is to test the hypothesis that rhythm perception abilities predict acoustic–prosodic entrainment, and this, in turn, contributes to conversational success. The overarching representation of the conceptual model can be found in Figure 1. As discussed previously, the hypothesized model is informed by theory, which suggests a relationship between rhythm perception and acoustic–prosodic entrainment (e.g., Phillips-Silver et al., 2010; Todd et al., 2002) and between acoustic–prosodic entrainment and conversational success (e.g., Local, 2007; Wilson & Wilson, 2005). The direction of the pathways was further established by temporal sequencing, which dictates that rhythm perception abilities must precede acoustic–prosodic entrainment, which, in turn, must precede conversational success.

To test this model, we focus specifically on speaking rate entrainment as our metric of entrainment and conversational quality as our metric of conversational success. We chose to focus on speaking rate entrainment because of robust evidence of its presence in conversations, as well as its association with conversational success (e.g., Manson et al., 2013; Street, 1984; Wynn & Borrie, 2020b). Additionally, studies showing lower levels of entrainment in populations with rhythm perception deficits are often specifically focused on speaking rate entrainment (e.g., Freeman & Pisoni, 2017; Wynn et al., 2018). We acknowledge that there are many ways to define and measure conversational success. Because of the broader implications of this work with clinical populations, our measure of conversational success was selected in consensus with an expert panel of speech-language pathologists (SLPs). When asked what conversational success meant, the majority of SLPs provided responses that focused on the "quality" of the conversation (e.g., high level of engagement, conversation feels natural, and conversation flows). Accordingly, in line with these responses, we use a measure of conversational quality based on third-party listener observations.

In our first objective, we focus on the relationship between rhythm perception abilities and speaking rate entrainment. Specifically, we compare a rhythmic beat discrimination task to a measure of static local synchrony (see Wynn & Borrie, 2020a, for a discussion of entrainment types) of speaking rate.[1] In order to understand the role of rhythm perception, relative to other factors that have been included in existing literature, we also test gender, partner familiarity, and social competence within our analyses. In our second objective, we examine the relationship between rhythm perception abilities and conversational quality. Specifically, we examine the direct pathway between rhythm perception and conversational quality, as well as an indirect pathway in which the relationship between these two factors is mediated by speaking rate entrainment.

We employ several methods of control to eliminate potential confounds that may affect the results of our model testing. First, we apply statistical control by including three disparate and broad control variables (i.e., gender, partner familiarity, and social competence) to account

---

[1]In line with the study of Yorkston et al. (2017) and their discussion on "speech versus speaking," we use the term *speaking rate* rather than *speech rate to* denote that information was collected during the active process of communication (i.e., dialogue) rather than a task devoid of interaction (e.g., shadowing task with prerecorded stimuli).

for the potential confounding effects these variables and their antecedents may have on established relationships. Second, we control for other confounding variables within our experimental design by using a highly controlled environment (i.e., controlled conversational task, same time, and location for all participants) and a well-controlled group of participants (i.e., college students of a similar age who are native English speakers). Additionally, we use a data collection methodology in which participants engage in multiple conversations in a round-robin–type sampling paradigm. This between–within design enables us to investigate the speaking rate entrainment patterns of each individual while accounting for the possibility that an individual may vary their level of entrainment, depending on their conversation partner.

Method

Collection of data for this study was part of a larger study examining entrainment in neurotypical adults using a round-robin data collection procedure. Within this present study, there were two phases of data collection: conversational collection and conversational rating. The conversation collection phase included a dialogue task, performed by a group of participants at the same time, as well as additional assessments, performed by each participant at individual testing sessions. The conversation rating phase was completed by a second set of participants using an online crowdsourcing website. Both phases are described in detail below.

Conversation Collection Phase

Participants

Data were collected from 20 neurotypical adults. All participants were undergraduate university students between the ages of 18 and 23 years ($M = 20.63$, $SD = 1.64$). Participants were native speakers of American English and passed a hearing screening administered at 20 dB for 1000, 2000, and 4000 Hz. Participants also reported no cognitive, speech, or language impairment. Participants were divided into two distinct experimental groups based on gender: one female group ($n = 10$) and one male group ($n = 10$). This grouping variable allowed us to consider the role of gender composition on acoustic–prosodic entrainment.

Dialogue Task

Ten participants of the same gender completed the conversation portion of the experiment at the same time. During the dialogue task, participants were divided into dyads and assigned a private room with a research assistant. Conversational dyads were seated facing one another

and were each fitted with a microphone connected to a single Zoom H4N portable digital recorder. Separate audio channels for each communication partner and standard settings (48 kHz; 16-bit sampling rate) were employed for audio recording of the dialogue task.

Dyads then completed the spoken dialogue Diapix task (Baker & Hazan, 2011; see also the study of Van Engen et al., 2010, for the original version). The task is a collaborative "spot-the-difference" task, used previously to study acoustic–prosodic entrainment (e.g., Borrie et al., 2015, 2019). In this task, each partner is given one of a pair of pictures and instructed not to show one another. Pictures are virtually identical but have 12 differences in detail between the two (e.g., three vs. four blankets). Participants were asked to communicate with one another to verbally identify as many differences as they could within a 5-min timeframe. Dyads were then free to verbally interact in any way they saw fit to problem-solve the task. The conversation ended when participants found all 12 differences or after 5 min had elapsed (whichever came first). After the task was completed, participants each answered a multiple-choice question, indicating the level of partner familiarity. Specifically, participants indicated if they had never met or were acquaintances, okay friends, or really good friends with their communication partner. Due to the small number of responses for some items and to allow for adequate group information to estimate differences, responses representing any degree of familiarity were combined into a single grouping variable (familiar vs. unfamiliar). Thus, our metric of partner familiarity reflects whether participants did or did not know their communication partner.[2] After answering this question, participants were divided into new dyad pairings and completed the same task with a new picture set. This process was repeated until each participant had completed the task with all nine other group members (and nine different picture sets). In total, 45 conversations were collected from each group (i.e., male group and female group), yielding a total of 90 conversations.

Additional Assessments

Rhythm perception abilities were assessed using the rhythm subtest of the Musical Ear Test (MET; Wallentin et al., 2010). This measure was selected because of its high validity as a measure of rhythm perception with essentially no floor or ceiling effects. The test uses a forced-choice

---

[2]In 88 of the 90 dyad pairings, partners' responses regarding the familiar versus. unfamiliar distinction were identical. In the two instances where responses differed, because entrainment was analyzed by participant (rather than by dyad), each participant's score reflected their perception of partner familiarity (i.e., their own response to the question).

paradigm for which participants must judge whether 52 pairs of rhythmic phrases are the same or different. The rhythm sequences within each phrase contain four to 11 woodblock beats and have a duration of one measure played at 100 beats/min. When phrases are different, they differ from one another by a single rhythmic change (for further details on how rhythm complexity is varied across phrase pairs, see Wallentin et al., 2010). No feedback regarding judgment accuracy is given at the time of testing. Scores for this assessment were calculated as the percentage of correctly answered items.

Social competence was assessed using the social acceptance subscale of the Self-Perception Profile for College Students (Harter, 2012). This subscale contains four structured alternative questions dealing with participants' perceptions of their social skills and ability to make friends. For each question, participants were asked to read two contrasting statements (e.g., "Some students like the way they interact with other people. Other students wish their interactions with other people were different") and identify which statement best describes them. The participant then decides whether the description they chose is "really true for me" or "sort of true for me." Each item is given a numeric value ranging from 1 to 4, with 1 representing low social competence and 4 representing high social competence. A total score for this measure was obtained by summing scores from each of the four questions.

Conversation Rating Phase

Listener Participants

Rating data were collected from 94 participants (male = 56, female = 38), ranging in age from 22 to 65 years (M = 37.72, SD = 10.57). Participants were recruited using Amazon Mechanical Turk (MTurk; http://www.mturk.com), an online crowdsourcing website successfully used in social science research involving listener ratings of speech recordings (e.g., Byun et al., 2015; Parker & Borrie, 2018; Yoho et al., 2019). Our inclusion criteria required that participants were native speakers of American English. Additionally, participants were required to meet the following qualifications on MTurk: (a) location confirmed in the United States, (b) Human Intelligence Task (HIT) approval rating of 99% or better (rating provided by MTurk to characterize worker performance), and (c) completion of a minimum of 10,000 HITs.

Stimuli

Audio clips that were approximately 60 s in length were extracted from each of the 90 dialogues elicited during the conversational collection portion of the experiment as described previously. This was done to maintain participant engagement with the task (i.e., so that participants

did not become bored and inattentive from listening to the entire duration of the conversation). As is frequently done in studies employing interactional ratings, the second minute of the conversation was used as a stimulus (e.g., Balaam et al., 2011; Bernieri et al., 1994; Ingham et al., 2001). In order to improve naturalness of the stimulus, each audio clip was created at the most natural breaking point within 5 s of the start and end of the 2-min mark of the conversation.

Rating Task

After obtaining informed consent, participants were instructed to put on headphones and check the sound on their computer. Participants were told they would be listening to audio clips and were given a short description of the dialogue task that had formerly been completed. However, they were not told the purpose of the study or given any information about the dyads. Participants were told that immediately after listening to each clip, they would use a 5-point Likert-type rating scale (1 = *disagree*, 3 = *neutral*, 5 = *agree*) to indicate the extent to which they agreed with the following statement: "*This conversation seemed to flow well (i.e., conversational partners were in-sync with each other)*.[3] Each participant was then presented with a randomly selected audio clip from the 90 total conversations. Participants selected a *play* button when they were ready to start each recording, and they were then required to listen to the entire duration of the audio clip before being allowed to make their rating and move to the next audio clip. This process continued until the participant had listened to and rated 15 audio clips.

Several measures were taken to better ensure high-quality data (e.g., data not collected from bots, inattention from research participants; Aguinis et al., 2020; Jonell et al., 2020; Keith et al., 2017) These included (a) two forced-response fill-in-the-blank questions requiring a good grasp of American English (e.g., "What is the punctuation mark commonly used at the end of a sentence?") after one-third and two-thirds of the task had been completed; (b) a financial bonus promised to participants who showed evidence that they had fully engaged in the task (granted when participants answered the fill-in-the-blank questions correctly); (c) disabling research participants from completing the study more than 1 time; and

---

[3]To verify the appropriateness of this question, we asked the panel of SLPs if they thought this question adequately captured conversational success. All 10 SLPs affirmed that this question was appropriate. Additionally, to ensure that the wording of the question appropriately encapsulated the idea of conversational quality, we asked 10 individuals (i.e., non-SLPs) how they interpreted this question. Responses closely matched the ideas of conversational quality expressed by our expert panel (e.g., *mutual engagement, neither person dominates, on the same page*).

(d) keeping audio samples short and limiting entire task to 25 min to maintain attention. Only one participant answered the fill-in-the-blank questions incorrectly, and their data were subsequently removed from analysis.

Acoustic Analysis

Trained research assistants manually coded each audio file, annotating and transcribing individual spoken utterances using the Praat textgrid function (Boersma & Weenink, 2020). Utterances were defined as a unit of speech with pauses no greater than 50 ms (Levitan & Hirschberg, 2011). Illustrations of the associated spectrograms were used to aid coding accuracy. These textgrids were used to calculate the number of syllables using an automated syllabification script from the Penn Phonetics Toolkit (Tauberer, 2008). This script relies on a prespecified list of vowels that are used to identify syllable nuclei. Onsets and codas for each syllable are then determined using a series of linguistic rules. In order to ensure accuracy of the automated syllable counting algorithm, syllables were counted by hand by a research assistant for 10% of the total data set (nine randomly selected conversations). Comparison indicated high agreement between the automated script and the research assistant, with a Pearson correlation r score of .99.

Speaking rate was then calculated by dividing the number of syllables by the total time for each utterance. In order to improve the accuracy of speaking rate measurement, utterances that were only one word long were removed from analysis. Due to the nature of the tasks, conversations were all of varying lengths. However, all conversations lasted at least 180 s. To ensure valid comparisons across participants and conversations, only data extracted from the first 180 s of each conversation were included in analysis.

Statistical Analysis

Examination of Entrainment

Our first research objective focused on the factors that predicted speaking rate entrainment (*a* pathways). For this objective, we used a measure of static local synchrony, operationally defined as simple, turn-by-turn alignment where one speaker's behavior in a single spoken utterance predicts the behavior of their communication partner in the subsequent adjacent spoken utterance. This was accomplished using linear mixed-effect models, similar to those used in previous studies (Borrie, Wynn, et al., 2020; Seidl et al., 2018). In these models, we assessed the degree to which the speaking rate of one speaker (leading rate) predicts the speaking rate of their conversation partner on a subsequent adjacent turn (subsequent rate) while accounting for individual variability

across participants and conversations. This model can be expressed as follows:

$$Y_{ct} = \beta_0 + \beta_1 \times Y_{ct-1} + \alpha_c + \alpha_i + \alpha_p \qquad (1)$$
$$+ \sum_j \beta_j \times covariates_i + \varepsilon_{ct},$$

$$\alpha_c \sim N\left(\mu_c, \sigma_c^2\right), \qquad (2)$$

$$\alpha_p \sim N\left(\mu_p, \sigma_p^2\right), \qquad (3)$$

$$\alpha_i \sim N\left(\mu_i, \sigma_i^2\right), \qquad (4)$$

$$\epsilon_{ct} \sim N\left(0, \sigma^2\right), \qquad (5)$$

where c is the cth conversation, i is the ith individual, p is the pth partner, Y is the speaking rate, t is the current time point (subsequent rate), and t-1 is the leading rate. Parameters α_c α_i, and α_p are the random intercepts, by conversation ID, individual ID, and partner ID, respectively.[4] The covariates included gender, partner familiarity, rhythm perception, and social competence. The degree to which an interlocutor's subsequent rate was predicted by their partner's leading rate (β_1) is indicative of the degree to which the interlocutor aligned their speaking rate to that of their communication partner. As such, only adjacent utterances between communication partners (as opposed to consecutive utterances produced by the same individual) were included. Notably, all continuous variables were mean centered before analysis.

In order to more fully establish whether entrainment patterns from the conversational corpus were meaningful and not capturing accidental or coincidental phenomena, we also analyzed the entrainment patterns of a sham corpus following an approach established in prior work (Borrie, Barrett, et al., 2019, 2020; Duran & Fusaroli, 2017). This was done by randomly lining up each speaker with partners not currently in conversation with them. For example, we used data from Speaker 1 when conversing with Speaker 2 and paired that up with data from Speaker 8 when Speaker 8 was conversing with Speaker 3. That is, Speaker 1 and Speaker 8 are now lined up in a "sham" conversation—one that never actually occurred. Thus, the sham corpus had all interdependent behavior removed while maintaining other aspects of conversation (e.g., acoustic–prosodic information of two speakers). This

---

[4]Random slopes for leading rate (Y_{ct-1}) by conversation ID, participant ID, and partner ID were also included within original models. However, model estimations were generally singular suggesting overspecification and were thus removed from the final models. Notably, results from these singular models did not lead to different conclusions than those presented herein. Further details regarding all tested models and results can be found at https://osf.io/zpxrm/.

was done for $n = 180$ sham conversations (i.e., one for each participant in each conversation). Using the same model specification as in the real dialogues, analyses assessed if entrainment patterns were detectable within sham conversations.

Entrainment and Predictive Factors

After entrainment patterns were confirmed within our conversational corpus, we used maximum-likelihood ratio tests to assess the degree to which four predictive factors predict entrainment (a pathways). Our four predictive factors were rhythm perception ability (continuous variable from rhythm subtest of the MET), participant gender (dichotomous variable of male vs. female), partner familiarity (dichotomous variable of knowing communication partner vs. not knowing communication partner), and social competence (continuous variable from social acceptance subscale of the Self-Perception Profile for College Students). In the first model (Model 1), subsequent rate was predicted by leading rate while controlling for gender, partner familiarity, rhythm perception ability, and social competence for the speaker of subsequent rate (i.e., the person entraining). This baseline model was then compared to a series of four models, to assess moderation of each predictive factor. Each model included an interaction between leading rate and one of the four predictive factors—rhythm perception (Model 2a), gender (Model 2b), social competence (Model 2c), and partner familiarity (Model 2d). Each model included the remaining predictive factors as control variables and included random intercepts for participant ID, partner ID, and conversation ID to account for intra-individual, intrapartner, and intraconversation variability. Thus, the general model for this objective can be expressed as follows:

$$Y_{ct} = \beta_0 + A_1 \times Y_{ct-1} + A_2 \times Y_{ct-1} \times factor + \alpha_c + \alpha_i + \alpha_p + \sum_j \beta_j \times covariates_i + \varepsilon_{ct}, \tag{6}$$

$$\alpha_c \sim N\left(\mu_c, \sigma_c^2\right), \tag{7}$$

$$\alpha_p \sim N\left(\mu_p, \sigma_p^2\right), \tag{8}$$

$$\alpha_i \sim N\left(\mu_i, \sigma_i^2\right), \tag{9}$$

$$\epsilon_{ct} \sim N\left(0, \sigma^2\right), \tag{10}$$

where *factor* is the corresponding predictive factor for that model, *covariates* include all predictive factors (including the one in the interaction), and A_2 is the estimate of interest. The investigation of any significant interactions included assessment of simple slopes to estimate entrainment by the levels of the predictive factors.

Relationship to Conversational Quality

Our second research objective examined the relationship between rhythm perception, speaking rate entrainment, and conversational quality. For this analysis, rhythm perception and conversational quality are a single measure per individual in each conversation. As such, it was necessary to also establish an entrainment value at that same level. To do this, we estimated the slope (β_1) of the line formed by regressing subsequent rate onto leading rate across adjacent conversational turns for each interlocutor in each dialogue. This was done using a series of linear regression models to obtain this aggregated measure for each individual in each conversation. This approach can be expressed as follows:

$$Y_{ct} = \beta_0 + \beta_1 \times Y_{ct-1}\epsilon_{ct} \tag{11}$$

for $m = 180$ total models (i.e., one model per speaker per conversation). Thus, the estimate (β_1) for each individual in each conversation (the entrainment score) represents the degree to which an individual's speaking rate was predicted by the rate of their partner's previous adjacent turn.

As for conversational quality, this measure came from the ratings provided in the second phase of the study. Because each participant was presented with audio clips in a random fashion, some audio clips received more ratings than others. However, all clips were rated between 10 and 17 times ($M = 16.0$, $SD = 1.4$). We then used these values within linear mixed-effects models to explore the relationship between speaking rate entrainment and conversational quality (pathway b) and the direct pathway between rhythm perception and conversational quality (pathway c). Thus, the model for the second objective can be expressed as follows:

$$Quality_{ipr} = B_0 + (B_1 + B_{1r} + B_{1i}) \times entrainment_{ipr} + C_1 \times rhythm_{ipr} + \gamma_r + \gamma_i + \gamma_p + \sum_{j=2} C_j \times covariates_{ipr} + \lambda_{ipr}, \tag{12}$$

$$\begin{matrix} \gamma_r \\ B_{1r} \end{matrix} \sim N\left[\begin{matrix} \mu_{\gamma_r} \\ \mu_{B_{1r}} \end{matrix}, \left(\begin{matrix} \sigma_{\gamma_r}^2 & \rho_{\gamma_r B_{1r}} \\ \rho_{B_{1r}\gamma_r} & \sigma_{B_{1r}}^2 \end{matrix}\right)\right], \tag{13}$$

$$\gamma_p \sim N\left(\mu_p, \sigma_p^2\right), \tag{14}$$

$$\begin{matrix} \gamma_i \\ B_{1i} \end{matrix} \sim N\left[\begin{matrix} \mu_{\gamma_i} \\ \mu_{B_{1i}} \end{matrix}, \left(\begin{matrix} \sigma_{\gamma_i}^2 & \rho_{\gamma_i B_{1i}} \\ \rho_{B_{1i}\gamma_i} & \sigma_{B_{1i}}^2 \end{matrix}\right)\right], \tag{15}$$

$$\lambda_{ipr} \sim N\left(0, \sigma^2\right), \tag{16}$$

where r is the rth rater, B_1 is the effect of entrainment on conversational quality, and C_1 is the "direct effect" of

Table 1. Results of fixed effects for real and sham conversations.

Term	b	SE	df	t value	p value
Real conversations					
Intercept	4.40	0.10	21	45.09	< .001***
Leading rate	0.08	0.01	3845	5.86	< .001***
Sham conversations					
Intercept	4.40	0.10	21	44.40	< .001***
Leading rate	0.004	0.02	3224	0.27	.79

Note. SE = standard error; df = degrees of freedom.

***p < .001.

rhythm on conversational quality. As before, the model includes gender, social competence, and partner familiarity as control variables. The random-effects structure included random intercepts and slopes (for entrainment) by participant ID, partner ID, and rater ID.[5]

All analyses were performed in the R statistical environment (R Version 3.6.1; R Development Core Team, 2019). Data cleaning and visualization relied on the "tidyverse" packages (Wickham et al., 2019). Mixed-effects models were analyzed using the lme4 (lme4 package Version 1.1–19; Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. Simple slope analysis was performed using the interactions (Long, 2019) package. All p values reported in conjunction with estimates of effect are based on Satterthwaite approximation to degrees of freedom. Analysis code and model output associated with this work are available at the study repository hosted at https://osf.io/zpxrm/.

Results

Descriptive Data

In total, our data set consisted of 4,699 adjacent conversational turns. The number of analyzed turns per participant per conversation ranged from 15 to 40 ($M = 26.1$, $SD = 5.8$). Analyses described below included seven variables: leading rate ($M = 4.5$, $SD = 1.5$), subsequent rate ($M = 4.4$, $SD = 1.4$), rhythm perception ability ($M = 67.2$, $SD = 9.3$), gender (45 male dialogues, 45 female dialogues), partner familiarity (40 familiar with partner, 140 unfamiliar with partner), social competence ($M = 12.1$, $SD = 2.3$), and conversational ratings ($M = 4.2$, $SD = 1.0$).

---

[5]A random intercept was originally included for conversational ID. However, the model estimation was singular due to this variable not accounting for variability in the outcome and was thus removed from the model.
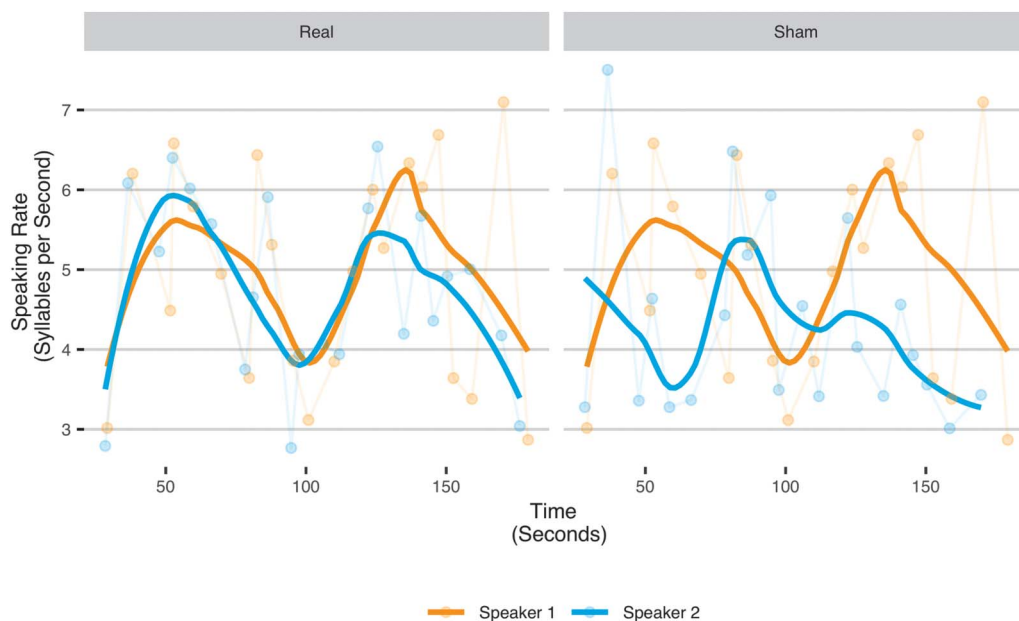
Examination of Entrainment

To investigate our first objective, the presence of acoustic–prosodic entrainment patterns in the speaking rate of conversational participants was first established. As shown in Table 1, findings from the real conversational corpus showed a significant level of entrainment ($p < .001$). Using this same analysis approach, we also analyzed the level of entrainment within the sham corpus. Contrastingly, findings from the sham conversational corpus showed no significant level of entrainment ($p = .79$). Additionally, the overall effect size of entrainment within our real conversational corpus ($b = 0.08$) was more than 20 times that of the sham corpus ($b = 0.004$). We thus conclude that, within the real conversational corpus, speakers are genuinely displaying a significant and meaningful level of entrainment. An illustrative example of the difference between real and sham data is presented in Figure 2. Using the data from one conversation, the speaking rate of Speaker 1 is shown with the preceding adjacent turns of Speaker 2 who is either their real conversation partner (left panel) or a sham partner (right panel). As illustrated, in real dialogues, partners aligned their speaking rate (i.e., as one increased, the other increased and as one decreased, the other decreased). This alignment pattern was not evident in the sham dialogues.

Examination of Predictive Factors

Likelihood ratio tests were performed to assess the best fitting model from a basic "main effects" model (Model 1) to more complex models that included potential moderators. Results indicated Model 2a (the model that includes an interaction between leading rate and rhythm perception) as the best-fitting model (see Table 2). Table 3 shows the estimates of Model 2a, revealing the significant interaction between rhythm perception and leading rate ($b = 0.004$, $p = .02$). Further tests for other potential moderators were also performed (i.e., likelihood ratio tests between Model 1 and each other models, tested individually). In each instance, Model 1 was revealed as the best-

Figure 2. Speaking rate of one speaker (Speaker 1) across a real and sham conversation with the preceding adjacent turns of their partner (Speaker 2). The left panel, illustrating the real conversation, shows speaking rate entrainment as supported by our analysis across the conversations. The right panel, illustrating the sham conversation, shows an absence of speaking rate entrainment.

fitting model (see Table 2). Thus, taken together, findings showed that entrainment was significantly influenced by rhythm perception but not by gender, partner familiarity, or social competence.

To further examine the relationship between entrainment and rhythm perception, a simple slope analysis was performed. This approach is designed to show the strength of the relationship in entrainment based on the level of a third variable—in this case, rhythm perception scores. Specifically, we assessed the strength of entrainment for when rhythm perception was 1 SD below the mean, at the mean, and 1 SD above the mean. As shown in Table 4, the slope for low rhythm perception scores was .05. The slope for medium rhythm perception scores was .08, and the slope

for high scores was .11. Thus, the degree of entrainment increased as rhythm perception scores increased. That is, the estimated effect size of entrainment was 60% higher for high rhythm perception scores relative to average scores and 120% greater for high rhythm perception scores relative to low rhythm perception scores.

Relationship to Conversation Quality

To investigate our second objective, additional mixed-effects models were used to examine the relationship between speaking rate entrainment, rhythm perception, and conversational quality. Because there was no significant relationship between acoustic–prosodic entrainment

Table 2. Linear mixed model fit indices for models of interest.

Model	AIC	BIC	Log likelihood	χ^2	χ^2 difference	df	p
Model 1	16,237	16,302	−8,109	16,217			
Model 2a	**16,233**	**16,304**	**−8,106**	**16,211**	**5.72**	**1**	**.02***
Model 2b	16,238	16,309	−8,108	16,216	0.58	1	.44
Model 2c	16,239	16,310	−8,109	16,217	0.006	1	.94
Model 2d	16,238	16,309	−8,108	16,216	1.37	1	.24

Note. Model 1 included main effects for leading rate and each of the four predictive factors. Each remaining model was compared to Model 1 on an individual basis. Model 2a included an interaction between leading rate and rhythm perception. Model 2b included an interaction between leading rate and gender. Model 2c included an interaction between leading rate and social competence. Model 2d included an interaction between leading rate and partner familiarity. AIC = Akaike information criterion; BIC = Bayesian information criterion; df = degrees of freedom. Boldface text represents best-fitting model.

*p < .05.

Table 3. Results of fixed effects for Model 2a of the analysis of predictive factors.

Term	b	SE	df	t value	p value
Intercept	4.12	0.11	23	36.99	< .001***
Leading rate	0.08	0.01	4038	5.63	< .001***
Rhythm perception	0.007	0.008	20	0.89	.39
Gender	0.53	0.16	22	3.42	.002**
Social competence	−0.01	0.03	20	−0.32	.75
Partner familiarity	0.10	0.06	77	1.71	.09
Leading Rate × Rhythm Perception	0.004	0.001	4575	2.40	.02*

Note. Only the interaction is indicative of a variable influencing entrainment of speaking rate. Here, the significance of gender simply denotes an overall difference in the speaking rates of male and female participants. *SE* = standard error; *df* = degrees of freedom.

*p < .05. **p < .01. ***p < .001.

and the other three factors, these were included as control variables but were not otherwise investigated within the analysis. We first examined the relationship between entrainment and conversational quality (pathway *b*). In our model (Model 3), conversational quality was predicted by entrainment score while controlling for rhythm perception, gender, partner familiarity, and social competence. Analysis of results showed no significant effect of entrainment on conversational quality (*b* = 0.20, *p* = .19). We used the same model to examine the direct effect of rhythm perception on conversational quality (pathway *c*). Again, findings showed no direct effect of rhythm perception on conversational quality (*b* = 0.001, *p* = .76).

Follow-up Analysis

Given the extensive literature showing a relationship between conversational entrainment and success, our findings regarding the relationship between conversational success (specifically conversational quality) and entrainment were unexpected. However, we note that examination of our model showed a high degree of within-speaker variability in the relationship between entrainment and conversational quality (intraclass correlation = .16). Therefore, it was quite possible that some of this heterogeneity could be predicted by other factors that were influencing this relationship, warranting a more fine-grained analysis. One potentially important factor that our data allowed us to account for was partner familiarity. We explored this relationship in a follow-up analysis using a linear mixed model (Model 4). As above, our model included conversational quality as the dependent variable and rhythm perception, gender, and social competence as control variables. The random-effects structure also included random intercepts and slopes (for entrainment) by participant ID, partner ID, and rater ID. However, this model included an interaction between partner familiarity and entrainment. Likelihood ratio tests between Model 3 (i.e., the original model) and Model 4 (i.e., the model with the interaction) revealed Model 4 as the best-fitting model.

As shown in Table 5, results revealed a significant interaction between entrainment and partner familiarity (*b* = 0.42, *p* = .03). Simple slopes analysis showed a relationship between entrainment and conversational quality in conversations between unfamiliar partners (*b* = 0.31, *p* = .056).[6] However, there was no effect of entrainment on conversational quality (*b* = −0.12, *p* = .58) in conversations between familiar partners.[7]

Discussion

The purpose of this study was to examine the role of rhythm perception abilities, firstly on speaking rate entrainment and secondly on the functional outcome of conversational quality. Findings from our first objective showed that rhythm perception abilities were indeed significantly predictive of speaking rate entrainment. Individuals with high rhythm perception abilities modified their speaking rate to align with their interlocutor more than individuals with low rhythm perception abilities. Thus, we provide empirical evidence to support the theoretical model proposed by Phillips-Silver et al. (2010)

---

[6]While this *p* value falls just short of the traditionally accepted threshold for statistical significance, there is no meaningful difference between *p* = .05 and *p* = .056. In consideration of this, taken together with results of likelihood ratio tests (showing the significant interaction), we can infer that a relationship likely exists in this instance. See the study of Wasserstein and Lazar (2016) for a wider discussion of *p* values.

[7]Given these findings, we ran an additional the given the model to investigate whether the direct effect of rhythm perception on conversational quality (pathway *c*) was moderated by partner familiarity. This model was similar to Model 4, but included an interaction term between rhythm perception and partner familiarity while controlling for entrainment, gender, and social competence. As with Model 4, the random-effects structure included random intercepts and slopes by participant ID, rater ID, and partner ID. Findings showed no significant interaction between rhythm perception and partner familiarity (*b* = −0.004, *p* = .34).

Table 4. Results of simple slopes analysis for Model 2a of the predictive factor analysis.

Group	Rhythm perception score	b	SE	95% confidence intervals	t value	p value
High rhythm perception	76.63	0.11	0.02	.07–.15	5.74	< .01
Average rhythm perception	67.43	0.08	0.01	.05–.11	5.63	< .01
Low rhythm perception	58.24	0.05	0.02	.01–.08	2.31	.02

Note. High rhythm perception = scores at 1 *SD* above the mean. Average rhythm perception = scores at the mean. Low rhythm perception = scores at 1 *SD* below the mean. *SE* = standard error.

specifying rhythm perception abilities as a component of acoustic–prosodic entrainment (here, specifically speaking rate entrainment). A large body of existing research has evidenced rhythmic cues as an important factor in aspects of speech communication, including word segmentation, speech recognition, and syntactic and lexical processing (e.g., Cutler & Norris, 1988; Dilley & McAuley, 2008; Mattys et al., 2005; Rothermich et al., 2012; Schmidt-Kassow & Kotz, 2008). Furthermore, studies have revealed that rhythm cues convey unspoken information, bind turn-taking dynamics, and express emotion (Couper-Kuhlen, 1993; Quinto et al., 2013; Wilson & Wilson, 2005). Here, we add to the existing body of literature, extending the role of rhythm in speech communication to the phenomenon of speaking rate entrainment.

While rhythm perception abilities were significantly predictive of speaking rate entrainment, gender composition, social competence, and partner familiarity were not. Within current literature, consideration of the role that these factors play in entrainment has led to largely disparate and inconclusive findings (e.g., Namy et al., 2002; Pardo, 2006; Pardo et al., 2012, 2018; Street & Cappella, 1989; Street & Murphy, 1987; Weidman et al., 2016; Weise et al., 2019; Yu et al., 2013). Discrepancies may be explained, in part, by considerable variability across experimental design and analysis. Thus, our findings, taken with past literature, lead us to conclude that while these factors may impact acoustic–prosodic entrainment, their influence is likely constrained by contexts and situations.

That rhythm perception facilitates acoustic–prosodic entrainment becomes more meaningful insofar as entrainment leads to functional outcomes in conversation. Our initial results showed no relationship between conversational quality and speaking rate entrainment. At first glance, these findings appear to contradict previous research that shows evidence of entrainment's role on conversational success (e.g., Manson et al., 2013; Michalsky et al., 2018; Polyanskaya et al., 2019a). However, given the large degree of within-participant variability in the relationship between entrainment and success, we chose to investigate this relationship further. Our findings indicated that the relationship between speaking rate entrainment and conversational quality was significantly influenced by partner familiarity. When partners were unfamiliar with one another, there was an effect of speaking rate entrainment on conversational quality. However, when partners were familiar with each other, the findings suggest that they relied less on entrainment of speaking rate to achieve this same measure of success. As the majority of acoustic–prosodic entrainment research has focused on conversations between unfamiliar partners, these findings appear to be aligned with previous literature.

Taken together, our findings suggest that familiar conversational partners are able to utilize other strategies to achieve conversational success that may not be readily available and/or appropriate to unfamiliar partners (e.g., shared experience and humor). These findings, while not originally anticipated, offer important insights into

Table 5. Results of fixed effects for Model 4 of the analysis of predictive factors.

Term	b	SE	df	t value	p value
Intercept	4.12	0.07	72	55.43	< .001***
Entrainment	0.31	0.15	20	2.03	.06
Rhythm perception	< 0.001	0.003	18	0.24	.81
Gender	0.07	0.07	24	0.91	.37
Social competence	0.01	0.01	18	1.05	.31
Partner familiarity	0.13	0.04	937	3.02	.003**
Entrainment × Partner Familiarity	0.42	0.20	126	−2.13	.03*

Note. *SE* = standard error; *df* = degrees of freedom.

*p < .05. **p < .01. ***p < .001.

conversational dynamics. Here, we lend support to the idea of a synergistic model of conversation (Dideriksen et al., 2019; Fusaroli et al., 2014). That is, successful conversations are not achieved through the use of a single fixed and stagnant strategy. Rather, conversational success may be attained via a number of different pathways, and dyads dynamically adjust and alter their approach depending on what strategies are available, appropriate, and convenient to them in a given conversational context.

Focusing on the findings from unfamiliar partners, we show that that rhythm perception abilities indirectly predicted conversational success, as measured by third-party listener observations of conversational quality. Importantly, when controlling for speaking rate entrainment, the direct relationship between these two factors is nonsignificant (and not influenced by partner familiarity). Rather, as aforementioned, results point toward an indirect relationship mediated through entrainment of speaking rate. That is, rhythm perception predicted speaking rate entrainment, and entrainment, in turn, predicted conversational success in unfamiliar partners. As previously discussed, the link between entrainment and success is not new. Novel here is that, using this established relationship, we provide some evidence of a possible mechanism (i.e., entrainment) through which rhythm perception ability may elevate conversational outcomes. These findings offer potential explanations regarding conclusions of previous studies. For example, Loeb et al. (2021) found that rhythmic abilities were predictive of participants' self-reported success in interactional contexts (e.g., leading group discussions and interacting at parties). Although not directly investigated, the authors speculated that this relationship might be driven by entrainment. Our findings provide evidence to support this postulation. Furthermore, other studies have suggested that musical instruction positively influences interpersonal relations (Eerola & Eerola, 2013; Spychiger et al., 1995). Although there are numerous possible explanations for such conclusions, our findings suggest that entrainment, at least in speaking rate, may be responsible, in part, for these effects. That is, as individuals increase their rhythm perception abilities through musical training (e.g., Ireland et al., 2018; Slater et al., 2013), they increase their ability to entrain to their partners, leading to higher levels of conversational success and subsequently increased social benefits.

Given differences in neurotypical populations, our findings offer potential implications in speech-language pathology. Previous research has indicated that rhythm perception deficits are present in many disorders, including autism (e.g., Isenhower et al., 2012), fluency disorders (e.g., Wieland et al., 2015), hearing impairment (e.g., Stabej et al., 2012), aphasia (e.g., Zipse et al., 2014), developmental language disorders (e.g., Weinert, 1992), and traumatic brain injury (e.g., Léard-Schneider &

Lévêque, 2020). Furthermore, evidence from other bodies of literature indicates deficits in entrainment (specifically speaking rate entrainment; e.g., Freeman & Pisoni, 2017; Sawyer et al., 2017; Wynn et al., 2018) and reductions in metrics of conversational success (e.g., Bone et al., 2013; Horton et al., 2020; Keck et al., 2017; Werle et al., 2021) in many of these clinical populations. While these components have been studied in isolation, our findings support the development of a casual framework to collectively consider rhythm perception deficits, acoustic–prosodic entrainment, and conversational success within the context of the conversation. More specifically, our study lends empirical support for earlier postulations that the rhythm deficits present in many populations with communication disorders may impact human interaction in terms of acoustic–prosodic entrainment and, subsequently, successful conversation (Borrie & Liss, 2014). Developing a causal framework to work within would allow hypothesis-driven investigations into rhythmic skills as a target to increase levels of acoustic–prosodic entrainment and conversational success, at least with unfamiliar communication partners. Indeed, rhythmic interventions have been shown to be an effective way to target communication in general (Fujii & Wan, 2014) and interactional skills in particular (e.g., Ghasemtabar et al., 2015; Hidalgo et al., 2017; Nayak et al., 2000; Sharda et al., 2018) across many clinical populations. Therefore, it is possible that these types of interventions may also increase acoustic–prosodic entrainment, leading ultimately to higher levels of conversational success.

Limitations and Future Directions

Compared to other areas of research in social sciences, the effect sizes of current results are relatively small. However, it is important to contextualize these values. Indeed, the effect sizes reported within this study are typical compared to many studies within entrainment literature (e.g., Cohen-Priva et al., 2017; Ko et al., 2016; Lubold et al., 2019; Schweitzer & Lewandowski, 2013). Key here is that the overall effect size of entrainment within our conversational corpus was large relative to that exhibited by entrainment at the level of chance. We offer some explanations for the small effect sizes reported in this study and more broadly within entrainment research. Within our specific analysis of static local synchrony, a conversation in which the highest level of entrainment possible was attained would only result in a beta coefficient equal to one, with values either above or below this coefficient, indicating less entrainment. Accordingly, by its very nature, this type of analysis yields inherently small effect sizes. Beyond this, we note that, within the context of entrainment, *more* does not always equate with *better*.

A certain degree of entrainment does lead to conversational success. However, at extreme levels (which would be reflected in effect sizes close or equal to one), entrainment becomes obsessive mimicry (i.e., copycatting), and, therefore, may be perceived as being patronizing, condescending, or overfacilitative (Fusaroli et al., 2014; Giles & Smith, 1979). Thus, within the context of entrainment, excessively large effect sizes would likely be indicative of unsuccessful conversations. Studies investigating optimal levels of entrainment for conversational success would allow adequate effect sizes to be established.

Our conversational corpus consisted of 90 conversations, however, we acknowledge that these conversations only came from 20 individuals. While our data collection design, in which many conversations were elicited from each participant, afforded a rigorous means of controlling for external factors including effects of communication partner and intra-individual variability, future studies with larger numbers of participants should be used to substantiate these findings. Our study also focused exclusively on speaking rate entrainment, driven by extensive evidence of its presence in conversations and its role in fostering conversational success (e.g., Manson et al., 2013; Street, 1984; Wynn & Borrie, 2020b). Research has shown that, beyond speaking rate, our perception of rhythm relies on many different factors across phonological, articulatory, phonatory, and temporal aspects of speech (Barry et al., 2009; Cumming, 2010, 2011; Galves et al., 2002; He, 2012; Lee & Todd, 2004; Liss et al., 2009; Ramus et al., 1999; White & Mattys, 2007). Therefore, entrainment of many different features might be influenced by rhythm perception ability. However, this can only be determined by future studies that examine whether findings generalize to other speech features. Beyond these factors, research could also examine other metrics of conversational success. Here, we focused on conversational quality as determined by third-party listener ratings. However, the mediated relationship between rhythm and other conversational metrics such as the dyad's perception of the conversation or objective measures of efficiency could be investigated.

While we only looked at rhythm perception abilities, it is likely that rhythm production abilities also play a mechanistic role in acoustic–prosodic entrainment and conversational success. Beyond simply perceiving the rhythm cues of others, entrainment requires individuals to produce rhythmic cues, enabling integration of the rhythmic input they receive and their speech output (Phillips-Silver et al., 2010). Furthermore, Polyanskaya et al. (2019b) have argued that when speakers produce highly rhythmic speech, it increases the predictability of the incoming signal, allowing for easier cue detection and entrainment between communication partners. Indeed, research suggests that speakers naturally increase their speech rhythmicity when trying to facilitate entrainment

and that highly rhythmic speech leads to higher levels of entrainment (Cerda-Oñate et al., 2021). Additionally, Borrie and colleagues (Borrie et al., 2015; Borrie, Barrett, et al., 2020) have evidenced disruptions in entrainment and success in conversations involving individuals with dysarthria, a speech disorder characterized by rhythm production deficits. Thus, a most comprehensive causal model of acoustic–prosodic entrainment and functional outcomes would include rhythm production and perception.

Conclusions

This study explored the relationship between rhythm perception, speaking rate entrainment, and conversational quality. Here, we showed that rhythm perception abilities predicted the degree of speaking rate entrainment, and in turn, entrainment predicted the level of conversational quality as determined by third-party listener ratings. Thus, taken together, our findings reveal an indirect relationship between rhythm perception abilities and conversational quality as mediated through speaking rate entrainment. These findings offer important implications for understanding the conversations of neurotypical populations and provide empirical evidence toward a framework for considering interactional success in clinical populations with rhythm perception deficits.

Acknowledgments

References

Aguilar, L. J., Downey, G., Krauss, R. M., Pardo, J. S., Lane, S., & Bolger, N. (2016). A dyadic perspective on speech accommodation and social connection: Both partners' rejection sensitivity matters. *Journal of Personality, 84*(2), 165–177. https://doi.org/10.1111/jopy.12149

Aguinis, H., Villamor, I., & Ramani, R. S. (2020). MTurk Research: Review and recommendations. *Journal of Management, 47*(4), 823–837. https://doi.org/10.1177/0149206320969787

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods, 43*(3), 761–770. https://doi.org/10.3758/s13428-011-0075-y

Balaam, M., Fitzpatrick, G., Good, J., & Harris, E. (2011). Enhancing interactional synchrony with an ambient display. In *Proceedings of the SIGCHI Conference on Human Factors*

in Computing Systems (pp. 867–876). Association for Computing Machinery. https://doi.org/10.1145/1978942.1979070

Barry, W., Andreeva, B., & Koreman, J. (2009). Do rhythm measures reflect perceived rhythm? *Phonetica, 66*(1–2), 78–94. https://doi.org/10.1159/000208932

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bernieri, F. J., Davis, J. M., Rosenthal, R., & Knee, C. R. (1994). Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin, 20*(3), 303–311. https://doi.org/10.1177/0146167294203008

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1) [Computer software]. http://www.praat.org

Bone, D., Lee, C. C., Chaspari, T., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. (2013). Acoustic-prosodic, turn-taking, and language cues in child–psychologist interactions for varying social demand. In *Proceedings of the Annual Conference of the International Speech Communication Association* (pp. 2400–2404). INTERSPEECH.

Borrie, S. A., Barrett, T. S., Liss, J. M., & Berisha, V. (2020). Sync pending: Characterizing conversational entrainment in dysarthria using a multidimensional, clinically-informed approach. *Journal of Speech, Language, and Hearing Research, 63*(1), 83–94. https://doi.org/10.3389/fpsyg.2015.01187

Borrie, S. A., Barrett, T. S., Willi, M. M., & Berisha, V. (2019). Syncing up for a good conversation: A clinically meaningful methodology for capturing conversational entrainment in the speech domain. *Journal of Speech, Language, and Hearing Research, 62*(2), 283–296. https://doi.org/10.1044/2018_JSLHR-S-18-0210

Borrie, S. A., Lansford, K. L., & Barrett, T. S. (2017). Rhythm perception and its role in perception and learning of dysrhythmic speech. *Journal of Speech, Language, and Hearing Research, 60*(3), 561–570. https://doi.org/10.1044/2016_JSLHR-S-16-0094

Borrie, S. A., & Liss, J. M. (2014). Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research, 57*(3), 815–824. https://doi.org/10.1044/2014_JSLHR-S-13-0149

Borrie, S. A., Lubold, N., & Pon-Barry, H. (2015). Disordered speech disrupts conversational entrainment: A study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. *Frontiers in Psychology, 6*, 1187. https://doi.org/10.3389/fpsyg.2015.01187

Borrie, S. A., Wynn, C. J., Berisha, V., Lubold, N., Willi, M. M., Coelho, C. A., & Barrett, T. S. (2020). Conversational coordination of articulation responds to context: A clinical test case with traumatic brain injury. *Journal of Speech, Language, and Hearing Research, 63*(8), 2567–2577. https://doi.org/10.1044/2020_JSLHR-20-00104

Byun, T. M., Halpin, P. F., & Szeredi, D. (2015). Online crowd-sourcing for efficient rating of speech: A validation study. *Journal of Communication Disorders, 53*, 70–83. https://doi.org/10.1016/j.jcomdis.2014.11.003

Canette, L.-H., Bedoin, N., Lalitte, P., Bigand, E., & Tillmann, B. (2019). The regularity of rhythmic primes influences syntax processing in adults. *Auditory Perception & Cognition, 2*(3), 163–179. https://doi.org/10.1080/25742442.2020.1752080

Cerda-Oñate, K., Toledo, G., & Ordin, M. (2021). Speech rhythm convergence in a dyadic reading task. *Speech Communication, 131*, 1–12. https://doi.org/10.1016/j.specom.2021.04.003

Cohen-Priva, C. U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *Journal of the Acoustical Society of America, 141*(5), 2989–2996. https://doi.org/10.1121/1.4982199

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction*. John Benjamins. https://doi.org/10.1075/pbns.25

Cumming, R. E. (2010). The interdependence of tonal and durational cues in the perception of rhythmic groups. *Phonetica, 67*(4), 219–242. https://doi.org/10.1159/000324132

Cumming, R. E. (2011). The language-specific interdependence of tonal and durational cues in perceived rhythmicality. *Phonetica, 68*(1–2), 1–25. https://doi.org/10.1159/000327223

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 113–121. https://doi.org/10.1037/0096-1523.14.1.113

Dideriksen, C., Fusaroli, R., Tylén, K., Dingemanse, M., & Christiansen, M. H. (2019). *Contextualizing conversational strategies: Backchannel, repair and linguistic alignment in spontaneous and task-oriented conversations*. CogSci.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*(3), 294–311. https://doi.org/10.1016/j.jml.2008.06.006

Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PLOS ONE, 12*(6), e0178140. https://doi.org/10.1371/journal.pone.0178140

Eerola, P.-S., & Eerola, T. (2013). Extended music education enhances the quality of school life. *Music Education Research, 16*(1), 88–104. https://doi.org/10.1080/14613808.2013.829428

Freeman, V., & Pisoni, D. B. (2017). Speech rate, rate-matching, and intelligibility in early-implanted cochlear implant users. *The Journal of the Acoustical Society of America, 142*(2), 1043–1054. https://doi.org/10.1121/1.4998590

Fujii, S., & Schlaug, G. (2013). Corrigendum: The Harvard Beat Assessment Test (H-BAT): A battery for assessing beat perception and production and their dissociation. *Frontiers in Human Neuroscience, 7*, 771. https://doi.org/10.3389/fnhum.2014.00870

Fujii, S., & Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: The SEP Hypothesis. *Frontiers in Human Neuroscience, 8*, 777. https://doi.org/10.3389/fnhum.2014.00777

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology, 32*, 147–157. https://doi-org/10.1016/j.newideapsych.2013.03.005

Galves, A., Garcia, J., Duarte, D., & Galves, C. (2002). *Sonority as a basis for rhythmic class discrimination*. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody, Aix-en-Provençe, France*. Speech Prosody Special Interest Group.

Ghasemtabar, S. N., Hosseini, M., Fayyaz, I., Arab, S., Naghashian, H., & Poudineh, Z. (2015). Music therapy: An effective approach in improving social skills of children with autism. *Advanced Biomedical Research, 4*(1), 157. https://doi.org/10.4103/2277-9175.161584

Giles, H., & Smith, P. (1979). Accommodation theory: Optimal levels of convergence. In H. Giles & R. N. St. Clair (Eds.), *Language and social psychology* (pp. 45–65). Blackwell.

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar

skills in children. *Developmental Science, 18*(4), 635–644. https://doi.org/10.1111/desc.12230

Grahn, J. A., & Schuit, D. (2012). Individual differences in rhythmic ability: Behavioral and neuroimaging investigations. *Psychomusicology: Music, Mind, and Brain, 22*(2), 105–121. https://doi.org/10.1037/a0031188

Gregory, S. W., Jr., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology, 70*(6), 1231–1240. https://doi.org/10.1037/0022-3514.70.6.1231

Harter, S. (2012). *Manual for the self-perception profile for college students*. University of Denver.

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: A common rhythm. *Frontiers in Psychology, 4,* 566. https://doi.org/10.3389/fpsyg.2013.00566

He, L. (2012). Syllabic intensity variations as quantification of speech rhythm: Evidence from both L1 and L2. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of Speech Prosody* (pp. 466–469). International Speech Communication Association. https://doi.org/10.5167/UZH-127263

Hidalgo, C., Falk, S., & Schön, D. (2017). Speak on time! Effects of a musical rhythmic training on children with hearing loss. *Hearing Research, 351,* 11–18. https://doi.org/10.1016/j.heares.2017.05.006

Horton, S., Humby, K., & Jerosch-Herold, C. (2020). Development and preliminary validation of a patient-reported outcome measure for conversation partner schemes: The Conversation and Communication Questionnaire for people with aphasia (CCQA). *Aphasiology, 34*(9), 1112–1137. https://doi.org/10.1080/02687038.2020.1738160

Ingham, R. J., Sato, W., Finn, P., & Belknap, H. (2001). The modification of speech naturalness during rhythmic stimulation treatment of stuttering. *Journal of Speech, Language, and Hearing Research, 44*(4), 841–852. https://doi.org/10.1044/1092-4388(2001/066)

Ireland, K., Parker, A., Foster, N., & Penhune, V. (2018). Rhythm and melody tasks for school-aged children with and without musical training: Age-equivalent scores and reliability. *Frontiers in Psychology, 9,* 426. https://doi.org/10.3389/fpsyg.2018.00426

Isenhower, R. W., Marsh, K. L., Richardson, M. J., Helt, M., Schmidt, R. C., & Fein, D. (2012). Rhythmic bimanual coordination is impaired in young children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 6*(1), 25–31. https://doi.org/10.1016/j.rasd.2011.08.005

Jonell, P., Kucherenko, T., Torre, I., & Beskow, J. (2020). Can we trust online crowdworkers?: Comparing online and offline participants in a preference test of virtual agents. In *Proceedings of the 2020 IVA Conference on Intelligent Virtual Agents.* Association for Computing Machinery.

Keck, C. S., Creaghead, N. A., Turkstra, L. S., Vaughn, L. M., & Kelchner, L. N. (2017). Pragmatic skills after childhood traumatic brain injury: Parents' perspectives. *Journal of Communication Disorders, 69,* 106–118. https://doi.org/10.1016/j.jcomdis.2017.08.002

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Frontiers in Psychology, 8,* 1359. https://doi.org/10.3389/fpsyg.2017.01359

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology, 2*(1), 125–156. https://doi.org/10.1515/labphon.2011.004

Ko, E.-S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language, 43*(2), 284–309. https://doi.org/10.1017/S0305000915000203

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTestPackage: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lagrois, M. É., Palmer, C., & Peretz, I. (2019). Poor synchronization to musical beat generalizes to speech. *Brain Sciences, 9*(7), 157. https://doi.org/10.3390/brainsci9070157

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLOS ONE, 7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

Lee, C. S., & Todd, N. P. M. (2004). Towards an auditory account of speech rhythm: Application of a model of the auditory 'primal sketch' to two multi-language corpora. *Cognition, 93*(3), 225–254. https://doi.org/10.1016/j.cognition.2003.10.012

Lehnert-LeHouillier, H., Terrazas, S., & Sandoval, S. (2020). Prosodic entrainment in conversations of verbal children and teens on the autism spectrum. *Frontiers in Psychology, 11,* 582221. https://doi.org/10.3389/fpsyg.2020.582221

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of the Annual Conference of the International Speech Communication Association* (pp. 3081–3084). INTERSPEECH.

Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality and attention, *Frontiers in Communication, 4.* https://doi.org/10.3389/fcomm.2019.00018

Léard-Schneider, L., & Lévêque, Y. (2020). Perception of music and speech prosody after traumatic brain injury. *PsyArXiv.* https://doi.org/10.31234/osf.io/w7cbf

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language, and Hearing Research, 52*(5), 1334–1352. https://doi.org/10.1044/1092-4388(2009/08-0208)

Local, J. (2007). Phonetic detail and the organization of talk-in-interaction. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)* (pp. 6–10).

Loeb, D., Reed, J., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Tuned in: Musical rhythm and social skills in adults. *Psychology of Music, 49*(2), 273–286. https://doi.org/10.1177/0305735619850880

Long, J. A. (2019). *interactions: Comprehensive, user-friendly toolkit for probing interactions. R package version 1.1.3.* https://cran.r-project.org/package=interactions

Lubold, N., Borrie, S. A., Barrett, T. S., Willi, M. M., & Berisha, V. (2019). Do conversational partners entrain on articulatory precision? In *Proceedings of the Annual Conference of the International Speech Communication Association* (pp. 1931–1935). INTERSPEECH.

Magne, C., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and Language, 153–154,* 13–19. https://doi.org/10.1016/j.bandl.2016.01.001

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts

cooperation. *Evolution and Human Behavior, 34*(6), 419–426. https://doi.org/10.1016/j.evolhumbehav.2013.08.001

Michalsky, J., Schoormann, H., & Niebuhr, O. (2018). Conversational quality is affected by and reflected in prosodic entrainment. In *International Conference on Speech Prosody* (pp. 389–392).

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*(4), 422–432. https://doi.org/10.1177/026192702237958

Natale, M. (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. https://doi.org/10.1037/0022-3514.32.5.790

Nayak, S., Wheeler, B. L., Shiflett, S. C., & Agostinelli, S. (2000). Effect of music therapy on mood and social interaction among individuals with acute traumatic brain injury and stroke. *Rehabilitation Psychology, 45*(3), 274–283. https://doi.org/10.1037/0090-5550.45.3.274

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proceedings of ACL-08: HLT, Short Papers* (pp. 169–172). Association for Computational Linguistics.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*(1), 190–197. https://doi.org/10.1016/j.wocn.2011.10.001

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics, 69,* 1–11. https://doi.org/10.1016/j.wocn.2018.04.001

Parker, M. A., & Borrie, S. A. (2018). Judgments of intelligence and likability of young adult female speakers of American English: The influence of vocal fry and the surrounding acoustic-prosodic context. *Journal of Voice, 32*(5), 538–545. https://doi.org/10.1016/j.jvoice.2017.08.002

Phillips-Silver, J., Aktipis, C. A., & Bryant, A. G. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Perception: An Interdisciplinary Journal, 28*(1), 3–14. http://doi.org/10.1525/mp.2010.28.1.3

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2). https://doi.org/10.1017/S0140525X04000056

Polonenko, M. J., Giannantonio, S., Papsin, B. C., Marsella, P., & Gordon, K. A. (2017). Music perception improves in children with bilateral cochlear implants or bimodal devices. *Journal of the Acoustical Society of America, 141*(6), 4494–4507. https://doi.org/10.1121/1.4985123

Polyanskaya, L., Samuel, A. G., & Ordin, M. (2019a). Speech rhythm convergence as a social coalition signal. *Evolutionary Psychology, 17*(3), 147470491987933. https://doi.org/10.1177/1474704919879335

Polyanskaya, L., Samuel, A. G., & Ordin, M. (2019b). Regularity in speech rhythm as a social coalition signal. *Annals of the New York Academy of Sciences, 1453*(1), 153–165. https://doi.org/10.1111/nyas.14193

Quinto, L., Thompson, W. F., & Keating, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology, 4,* 184. https://doi.org/10.3389/fpsyg.2013.00184

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

R Development Core Team. (2019). *R: A language and environment for statistical computing.* https://www.R-project.org/

Reichel, U. D., Beňuš, Š., & Mády, K. (2018). Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication, 100,* 46–57. https://doi.org/10.1016/j.specom.2018.04.009

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia, 50*(2), 232–244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Sawyer, J., Matteson, C., Ou, H., & Nagase, T. (2017). The effects of parent-focused slow relaxed speech intervention on articulation rate, response time latency, and fluency in preschool children who stutter. *Journal of Speech, Language, and Hearing Research, 60*(4), 794–809. https://doi.org/10.1044/2016_JSLHR-S-16-0002

Schmidt-Kassow, M., & Kotz, S. A. (2008). Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Research, 1226,* 144–155. https://doi.org/10.1016/j.brainres.2008.06.017

Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. In F. Bimbot (Ed.), *Proceedings of the 14th Annual Conference of the International Speech Communication Association* (pp. 525–529). INTERSPEECH.

Schweitzer, A., Lewandowski, N., & Duran, D. (2017). Social attractiveness in dialogs. In *Proceedings of the Annual Conference of the International Speech Communication Association* (pp. 2243–2247). INTERSPEECH.

Seidl, A., Cristia, A., Soderstrom, M., Ko, E.-S., Abel, E. A., Kellerman, A., & Schwichtenberge, A. J. (2018). Infant-mother acoustic-prosodic alignment and developmental risk. *Journal of Speech, Language, and Hearing Research, 61*(6), 1369–1380. https://doi.org/10.1044/2018_JSLHR-S-17-0287

Sharda, M., Tuerk, C., Chowdhury, R., Jamey, K., Foster, N., Custo-Blanch, M., Tan, M., Nadig, A., & Hyde, K. (2018). Music improves social communication and auditory–motor connectivity in children with autism. *Translational Psychiatry, 8*(1), 231. https://doi.org/10.1038/s41398-018-0287-3

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians. *Cognitive Processing, 17*(1), 79–87. https://doi.org/10.1007/s10339-015-0740-7

Slater, J., Tierney, A., & Kraus, N. (2013). At-risk elementary school children with one year of classroom music instruction are better at keeping a beat. *PLOS ONE, 8*(10), e77250. https://doi.org/10.1371/journal.pone.0077250

Spychiger, M., Patry, J.-L., Lauper, G., Zimmerman, E., & Weber, E. (1995). Does more music teaching lead to a better social climate. In R. Olechowski & G. Svik (Eds.), *Experimental research in teaching and learning* (pp. 322–336). Peter Lang.

Stabej, K. K., Smid, L., Gros, A., Zargi, M., Kosir, A., & Vatovec, J. (2012). The music perception abilities of prelingually deaf children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology, 76*(10), 1392–1400. https://doi.org/10.1016/j.ijporl.2012.07.004

Street, R. L., Jr. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research, 11*(2), 139–169. https://doi.org/10.1111/j.1468-2958.1984.tb00043.x

Street, R. L., Jr. & Cappella, J. N. (1989). Social and linguistic factors influencing adaptation in children's speech. *Journal of*

Psycholinguistic Research, 18(5), 497–519. https://doi.org/10.1007/BF01067313

Street, R. L., Jr., & Murphy, T. (1987). Interpersonal orientation and speech behavior. *Communication Monographs, 54*(1), 42–62. https://doi.org/10.1080/03637758709390215

Tauberer, J. (2008). *P2TK automated syllabifier.* https://sourceforge.net/p/p2tk/code/HEAD/tree/python/syllabify/

Taylor, P. J., & Thomas, S. (2008). Linguistic style matching and negotiation outcome. *Negotiation & Conflict Management Research, 1*(3), 263–281. https://doi.org/10.1111/j.1750-4716.2008.00016.x

Todd, N. P. M., Lee, C. S., & O'Boyle, D. J. (2002). A sensori-motor theory of temporal tracking and beat induction. *Psychological Research, 66*(1), 26–39. https://doi.org/10.1007/s004260100071

Tordjman, S., Davlantis, K. S., Georgieff, N., Geoffray, M.-M., Speranza, M., Anderson, G. M., Xavier, J., Botbol, M., Oriol, C., Bellissant, E., Vernay-Leconte, J., Fougerou, C., Hespel, A., Tavenard, A., Cohen, D., Kermarrec, S., Coulon, N., Bonnot, O., & Dawson, G. (2015). Autism as a disorder of biological and behavioral rhythms: Toward new therapeutic perspectives. *Frontiers in Pediatrics, 3,* 1. https://doi.org/10.3389/fped.2015.00001

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The wildcat corpus of native and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech, 53*(1), 510–540. https://doi.org/10.1177/0023830910372495

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences, 20*(3), 188–196. https://doi.org/10.1016/j.lindif.2010.02.004

Wasserstein, R., & Lazar, N. A. (2016) The ASA statement on p-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Weidman, S., Breen, M., & Haydon, K. C. (2016). Prosodic speech entrainment in romantic relationships. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of Speech Prosody* (pp. 508–512). International Speech Communication Association.

Weinert, S. (1992). Deficits in acquiring language structure: The importance of using prosodic cues. *Applied Cognitive Psychology, 6*(6), 545–571. https://doi-org.10.1002/acp.2350060607

Weise, A., Levitan, S. I., Hirschberg, J., & Levitan, R. (2019). Individual differences in acoustic-prosodic entrainment in spoken dialogue. *Speech Communication, 115,* 78–87. https://doi-org.dist.lib.usu.edu/10.1016/j.specom.2019.10.007

Werle, D., Winters, K. L., & Byrd, C. T. (2021). Preliminary study of self-perceived communication competence amongst adults who do and do not stutter. *Journal of Fluency Disorders, 70,* 105848. https://doi.org/10.1016/j.jfludis.2021.105848

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501–522. https://doi.org/10.1016/j.wocn.2007.02.003

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhm, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *The Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wieland, E. A., McAuley, J. D., Dilley, L. C., & Chang, S.-E. (2015). Evidence for a rhythm perception deficit in children who stutter. *Brain and Language, 144,* 26–34. https://doi.org/10.1016/j.bandl.2015.03.008

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review, 12*(6), 957–968. https://doi.org/10.3758/BF03206432

Wynn, C. J., & Borrie, S. A. (2020a). *Classifying conversational-entrainment of speech behavior: An updated framework and review.* PsyArXiv. https://doi.org/10.31234/osf.io/nrj2z

Wynn, C. J., & Borrie, S. A. (2020b). Methodology matters: The impact of research design on conversational entrainment outcomes. *Journal of Speech, Language, and Hearing Research, 63*(5), 1352–1360. https://doi.org/10.1044/2020_JSLHR-19-00243

Wynn, C. J., Borrie, S. A., & Sellers, T. P. (2018). Speech rate entrainment in children and adults with and without autism spectrum disorder. *American Journal of Speech-Language Pathology, 27*(3), 965–974. https://doi.org/10.1044/2018_AJSLP-17-0134

Yoho, S. E., Borrie, S. A., Barrett, T. S., & Whittaker, D. (2019). Are there sex effects for speech intelligibility in American English? Examining the influence of talker, listener, and methodology. *Attention, Perception, & Psychophysics, 81*(2), 558–570. https://doi.org/10.3758/s13414-018-1635-3

Yorkston, K., Baylor, C., & Britton, D. (2017). Speech versus speaking: The experiences of people with Parkinson's disease and implications for intervention. *American Journal of Speech-Language Pathology, 26*(2S), 561–568. https://doi.org/10.1044/2017_ajslp-16-0087

Yu, A. C., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLOS ONE, 8*(9), e74746. https://doi.org/10.1371/journal.pone.0074746

Zipse, L., Worek, A., Guarino, A. J., & Shattuck-Hufnagel, S. (2014). Tapped out: Do people with aphasia have rhythm processing deficits? *Journal of Speech, Language, and Hearing Research, 57*(6), 2234–2245. https://doi.org/10.1044/2014_jslhr-l-13-0309