Sync Pending: Characterizing Conversational Entrainment in Dysarthria Using a Multidimensional, Clinically Informed Approach

Stephanie A. Borrie,[a] Tyson S. Barrett,[b] Julie M. Liss,[c] and Visar Berisha[c,d]

Purpose: Despite the import of conversational entrainment to successful spoken dialogue, the systematic characterization of this behavioral syncing phenomenon represents a critical gap in the field of speech pathology. The goal of this study was to acoustically characterize conversational entrainment in the context of dysarthria using a multidimensional approach previously validated in healthy populations (healthy conversations; Borrie, Barrett, Willi, & Berisha, 2019).
Method: A large corpus of goal-oriented conversations between participants with dysarthria and healthy participants (disordered conversations) was elicited using a "spot the difference" task. Expert clinical assessment of entrainment and a measure of conversational success (communicative efficiency) was obtained for each of the audio-recorded conversations. Conversational entrainment of acoustic features representing rhythmic, articulatory, and phonatory dimensions of speech was identified using cross-recurrence quantification analysis with clinically informed model parameters and validated with a sham condition involving conversational participants who did not converse with one another. The relationship between conversational entrainment and communicative efficiency was examined.

Results: Acoustic evidence of entrainment was observed in phonatory, but not rhythmic and articulatory, behavior, a finding that differs from healthy conversations in which entrainment was observed in all speech signal dimensions. This result, that disordered conversations showed less acoustic entrainment than healthy conversations, is corroborated by clinical assessment of entrainment in which the disordered conversations were rated, overall, as being less in sync than healthy conversations. Furthermore, acoustic entrainment was predictive of communicative efficiency, corroborated by a relationship between clinical assessment and the same outcome measure.
Conclusions: The findings confirm our hypothesis that the pathological speech production parameters of dysarthria disrupt the seemingly ubiquitous phenomenon of conversational entrainment, thus advancing entrainment deficits as an important variable in dysarthria, one that may have causative effects on the success of everyday communication. Results further reveal that while this approach provides a broad overview, methodologies for characterizing conversational entrainment in dysarthria must continue to be developed and refined, with a focus on clinical utility.
Supplemental Material: https://osf.io/ktg5q

[a]Department of Communicative Disorders and Deaf Education, Utah State University, Logan
[b]Department of Psychology, Utah State University, Logan
[c]Department of Speech and Hearing Science, Arizona State University, Tempe
[d]School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe
Correspondence to Stephanie A. Borrie: stephanie.borrie@usu.edu

Conversational entrainment, the communication phenomenon whereby communication partners synchronize their behavior, is considered key to productive and fulfilling communicative interactions. This relationship has been observed with entrainment of numerous verbal behaviors (e.g., Chartrand & Bargh, 1999; Lee et al., 2014). For example, studies have reported that task-oriented conversations characterized by high levels of acoustic-prosodic (Borrie, Lubold, & Pon-Barry, 2015) or lexical (Nenkova, Gravano, & Hirschberg, 2008) entrainment are also characterized by high levels of communicative efficiency and success in verbal problem-solving tasks. It has been advanced that synchronized behavior underlies tightly coupled production and comprehension processes and that this

coupling greatly reduces the computational load of language processing in spoken dialogue (Pickering & Garrod, 2004). Entrainment has also been associated with important pragmatic elements of conversation, including turn-taking, interaction smoothness, building rapport, fostering social bonds, and maintaining interpersonal relationships (e.g., Bailenson & Yee, 2005; Chartrand & Bargh, 1999; Wilson & Wilson, 2005; see Beňuš, 2014, for a review). Thus, it appears that entrainment serves cognitive and pragmatic functions essential for successful conversation. Lack of entrainment could, therefore, negatively impact conversational success, contributing to social isolation and diminished quality of life.

In order to entrain in the speech domain, individuals must be able to produce and perceive speech behavior (i.e., rhythm, articulation, phonation) and modify these speech behaviors to align more closely with their conversational partner (see Phillips-Silver, Aktipis, & Bryant, 2010, for a general overview). However, as is well established in the field of communication disorders, many individuals are unable to adequately produce, perceive, and/or modify speech behavior (see Borrie & Liss, 2014). Individuals with the neurological speech disorder of dysarthria, for example, have difficulty producing speech, whereas individuals with congenital or acquired hearing loss have difficulty perceiving speech. Ostensibly, the motor limitations and deviant speech production parameters that characterize dysarthria would negatively impact entrainment, as would the speech perception limitations induced by impaired hearing. There is certainly evidence that communication disorders such as dysarthria negatively impact presumed functions of conversational entrainment, including turn-taking timing (Comrie, Mackenzie, & McCall, 2001) and dealing with breakdowns in understanding (Bloch & Wilkinson, 2009), and a recent large-scale study revealed that, even after controlling for age, gender, health, and disability, communication impairment was a significant predictor for key aspects of social function, including fewer friends, reduced social participation, and decreased social self-efficacy (Palmer et al., 2019). While conversational entrainment has been studied in many disciplines, it surprisingly has not been systematically expanded to the field of communication disorders where its implications may have direct clinical relevance.

There is preliminary evidence of entrainment deficits in clinical populations including dysarthria (Borrie et al., 2015), hearing impairment (Freeman & Pisoni, 2017), autism spectrum disorder (Wynn, Borrie, & Sellers, 2018) fluency disorders (Sawyer, Matteson, Ou, & Nagase, 2017), and brain injury (Gordon, Rigon, & Duff, 2015). For example, using a small selection of basic acoustic-prosodic features (pitch, intensity, jitter, shimmer) and two simple local, turn-by-turn entrainment measures (synchrony and proximity; see also Levitan & Hirschberg, 2011), Borrie et al. (2015) observed that goal-oriented conversations between participants with dysarthria and healthy participants were characterized by significantly less entrainment than goal-oriented conversations involving two healthy participants. Furthermore, conversations characterized by lower levels of acoustic-prosodic entrainment, across both disordered and healthy populations, tracked with lower levels of communicative efficiency, quantified as success in achieving the goals of the spoken dialogue task.

While entrainment deficits may be real and of potential consequence in populations with communication disorders, the jump to explicit clinical implications has been, up until now, largely speculative. To date, speech signal entrainment has been measured as a purely objective phenomenon and correlated with measures that contribute to some aspect of conversational success (e.g., communicative efficiency). However, to establish the clinical value of conversational entrainment for application in speech-language pathology, it is necessary to link objective measures of speech signal entrainment with expert clinical assessment of synchronization. Furthermore, initial studies of entrainment in populations with communication disorders have focused on a few specific, easy-to-extract acoustic-prosodic features of interest (e.g., pitch, speech rate) and have largely treated entrainment as a local phenomenon in which behaviors influence one another on a turn-by-turn basis. That is, increases of pitch, for example, by one partner are immediately echoed by the other partner in the subsequent speaking turn. This early work is a valuable preliminary step, establishing that entrainment is an operational interaction process in the context of disordered speech (Borrie & Liss, 2014; Borrie et al., 2015). However, the next step must be a more comprehensive characterization of the communication phenomenon, across many acoustic features and multiple timescales.

With this in mind, we recently designed a computational methodology for characterizing conversational entrainment in the speech domain, with entrainment conceptualized as the emergent behavioral signal that captures interdependencies between conversational partners (also termed *interpersonal coordination*; Duran & Fusaroli, 2017). The design of this computational methodology was motivated by three key foundations. First, the methodology must be informed by real-world evidence of conversational entrainment, derived from expert clinical assessment of conversation. Second, the methodology should not self-select acoustic features of interest; rather, it must capture entrainment across features that represent multiple dimensions of the speech signal. Finally, the methodology must make no assumptions about who is leading the conversation or the temporal scale at which entrainment occurs. Indeed, as noted by Duran and Fusaroli (2017), "coordinated behaviors do not need to be isomorphic and occur close in time…rather, they can be distributed and loosely coupled across various local and global temporal scales" (p. 2). With these foundations set, we built a clinically informed method, pulling together a number of well-established techniques to capture entrainment, and validated the approach in a large conversational corpus of spoken dialogue in healthy populations (Borrie, Barrett, Willi, & Berisha, 2019). Using this methodology, we found evidence of entrainment across all acoustic feature sets, representing rhythmic, articulatory, and phonatory dimensions of speech. Methodology validation was established by statistical comparisons with a corpus of

randomly generated sham dialogues involving participants who did not converse with one another, and a series of predictive models of communicative efficiency. The creation of this clinically informed methodology for capturing entrainment in the speech domain has set the stage for characterization of conversational entrainment in populations with communication disorders.

The purpose of this study was to use the clinically informed methodology, developed and validated with healthy conversations, to acoustically characterize conversational entrainment in a corpus of experimentally elicited, goal-oriented conversations between individuals with dysarthria and healthy interlocuters. Toward this purpose, we addressed the following two key research questions: (a) To what extent, and on which speech signal dimensions, is there objective acoustic evidence of conversational entrainment? (b) Do acoustic measures of conversational entrainment predict a measure of conversational success (communicative efficiency)? Given dysarthria is characterized by speech production deficits, it was hypothesized that these conversations would show less objective evidence of acoustic entrainment than the healthy conversations in Borrie et al. (2019) and that acoustic entrainment would predict communicative efficiency. As a secondary exploration of the data, we also investigated potential relationships with and between expert clinical assessments of entrainment, communicative efficiency, and dysarthria severity ratings.

Method

Participants

This study is based on a corpus of experimentally elicited conversations from 52 dyads, involving 104 participants (64 females and 40 males) comprising 52 healthy (i.e., neurotypical) individuals and 52 individuals with dysarthria. The healthy participants were all native speakers of American English with no neurological history or presence of disordered speech patterns. The participants with dysarthria were also native speakers of American English but had a clinical diagnosis of a dysarthria as evaluated by a speech-language pathologist (SLP) not associated with this study. Dysarthrias of various etiologies were included in this initial characterization study; however, the majority of individuals had a diagnosis of Parkinson's disease. Dysarthria severity was largely mild ($n = 27$), but some individuals presented as mild-to-moderate ($n = 11$), moderate ($n = 10$), or severe ($n = 4$). Dysarthria severity ratings of 1 (*mild*) to 4 (*severe*) were based on perceptual estimates from the SLP who diagnosed the presence of dysarthria. Individuals with dysarthria were excluded from the study if concomitant impairments in speech (e.g., apraxia of speech) or language (e.g., aphasia) were identified. The composition of each dyad was controlled, so that each conversation involved one healthy participant and one participant with dysarthria. Gender, however, was not controlled for when forming dyads, so some dyads were female–female ($n = 19$), other dyads were female–male ($n = 26$), and others were male–male ($n = 7$).

Conversational Task

The audio recording procedure and conversational task were identical to that used for methodology development and validation in Borrie et al. (2019). Each dyad participated in a single recording session. Conversational partners were seated facing one another and fitted with wireless CVL Lavalier microphones, synced with a Shure BLX188 DUAL Lavalier System connected to a Zoom H4n Portable Digital Recorder. Separate audio channels for each conversational partner and standard settings (48 kHz; 16-bit sampling rate) were employed for audio recording of the conversational task.

The conversation task was based on the Diapix task, a collaborative "spot the difference" task whereby dyads must work together, verbally comparing scenes, to identify 10 differences between sets of pictures (Van Engen et al., 2010). Each partner in the dyad was given one of a pair of pictures and instructed to hold their picture at an angle at which it was not visible to their partner sitting across the table from them. The pair of pictures depicts virtually identical scenes (e.g., yard, beach), differing from one another by 10 small details (e.g., number of people, color of t-shirt). The dyad was told that their goal was to work together, simply by speaking to one another, to identify the 10 differences between the pair of pictures, and to do so as accurately and as quickly as possible. When all the differences were identified, the dyad was given another pair of pictures to work through. Dyads were tasked with working through as many pairs of pictures as possible in a 10-min timeframe. No additional rules (i.e., who could talk when) or roles (i.e., giver, receiver) were given, so dyads were free to verbally interact in any way they saw fit to problem-solve the task. Picture pair sets were presented in the same order across all dyads. It is important to acknowledge that the task is not considered cognitively demanding; however, conversational partners must work together to be successful (i.e., identify the differences).

Objective Measure of Communicative Efficiency

An objective measure of an aspect of conversational success, communicative efficiency, can be obtained from a dialogue elicited using the Diapix task. The Diapix task required the conversational partners to work together as accurately and quickly as possible to identify the 10 differences between a pair of pictures and then move on to the next picture pair. The communicative efficiency score is a tally of the total number of differences identified in the 10-min recording. The measure of communicative efficiency is, therefore, an objective measure of how proficiently the dyad used verbal communication to collaboratively work through the demands of the goal-oriented dialogue task. This objective measure is used to examine the predictive relationship between acoustic entrainment and conversational success—a relationship previously reported with other

acoustic measures of entrainment (e.g., Borrie et al., 2019, 2015; Willi, Borrie, Barrett, Tu, & Berisha, 2018).

Expert Clinical Assessment of Conversational Entrainment

Five SLPs evaluated the 52 conversations using a 7-point Likert-type rating scale (1 = *strongly disagree*, 4 = *neutral*, 7 = *strongly agree*) based on the extent that they agreed with the following statement: *The conversation pair sound like they are in-sync or aligned with one another. High ratings (scores above 4) should be indicative of a natural cohesiveness to the interaction, smooth turn-taking and conversational flow, and a sense of rapport and connection between conversational participants whereas low ratings (scores below 4) should be indicative of an awkward, disconnected, and disengaged interaction.* Thus, this score is indicative of a holistic impression of degree of conversational entrainment. As in Borrie et al. (2019), the clinicians were required to listen to the first 2 min[1] of a conversation before making their assessment rating. Interrater reliability between the five clinician judgments across the conversations was high (Cronbach's α = .9). This provides important validation that the SLPs in this study were reliable, with one another, at assessing conversational success as it relates to a holistic view of conversational entrainment. The ratings from five SLPs were averaged to achieve a mean expert clinical assessment score for each of the 52 conversations. As illustrated in Figure 1, over 70% of conversations were rated below a score of 4, which is in stark contrast to the healthy conversations in our previous study where over 70% were rated above 4. This holistic measure of entrainment is used to set the parameters for the objective analysis of acoustic entrainment.

Computational Methodology Overview

Figure 2 presents a high-level diagram of the clinically informed methodology for characterizing conversational entrainment in the acoustic domain. In brief, the methodology begins with extracting large acoustic feature sets that represent rhythmic, phonatory, and articulatory dimensions of speech. These feature sets are then reduced with independent component analysis (ICA), retaining shared information among the individual acoustic behaviors (Comon, 1992; Marchini, Heaton, & Ripley, 2017). Cross-recurrence quantification analysis (CRQA), a nonlinear technique that allows us to quantify shared organization of behavior over time (Coco & Dale, 2014; Zbilut, Giuliani, & Webber, 1998), with model parameters set by expert clinical assessment of the conversations, is then used to capture global acoustic entrainment in the speech domain. Sham conversations enable the identification of entrained versus not entrained measures (i.e., for methodology output to be considered entrained, the measures must be significantly different

from those produced by sham conversations). Several statistical models are then used to provide information regarding the predictive accuracy of acoustic entrainment on a presumed functional outcome of entrainment, communicative efficiency. Methodology components are described below, but for comprehensive details, please refer to Borrie et al. (2019).

Acoustic Feature Extraction

Trained research assistants manually coded each conversation audio file, annotating individual spoken utterances by speaker using the Praat TextGrid function (Boersma & Weenink, 2017). A spoken utterance is defined as a pause-free unit of speech, where pauses are greater than 50 ms, from a single speaker. Thus, pauses less than 50 ms are included in the utterance. This definition of a spoken utterance is the same as interpausal units (Levitan & Hirschberg, 2011). Audio files were then normalized using a reference level and down-sampled to 16 kHz prior to speech feature extraction.

Five large acoustic feature sets were extracted from each individual spoken utterance, including (a) envelope modulation spectrum (EMS; 60 features), (b) rhythm metrics (12 features), (c) long-term average spectrum (LTAS; 99 features), (d) mel-frequency cepstral coefficients (MFCCs; 234 features), and (e) voice report (24 features). These feature sets, which have been previously reported on (e.g., Berisha, Liss, Sandoval, Utianski, & Spanias, 2014; Tu, Berisha, & Liss, 2017; Tu, Jiao, Berisha, & Liss, 2016; Willi et al., 2018), are considered to reflect rhythmic (EMS, rhythm metrics), articulatory (LTAS, MFCC), and phonatory (voice report) dimensions of speech signal behavior (e.g., Cleveland, Sunberg, & Stone, 2001; Dellwo, Fourcin, & Abberton, 2013; Liss, LeGendre, & Lotto, 2010; also see Borrie et al., 2019, for dimension justification). For comprehensive details of feature calculation, please refer to Supplemental Material S1.

Acoustic Feature Reduction

As noted above, each speech feature set is made up of a number of features. We used ICA to perform feature reduction for the five feature sets: EMS (using the five features—of the total 60—relating to a center frequency of 480; this is likely to capture the rhythmic patterns associated with changes in vowel energy), rhythm metrics (using all 12 features), LTAS (using all 99 features), MFCC (using all 234 features), and voice report (using all 24 features). Each feature set had a high amount of shared variability (all Cronbach's α > .60), signifying that the ICA captured a high degree of the original variability across the features. Thus, the ICA produced five variables, representing rhythmic (EMS, rhythm metrics), articulatory (LTAS, MFCC), and phonatory (voice report) speech dimensions, to be used in objective acoustic entrainment analyses.

Objective Acoustic Entrainment: CRQA

Acoustic entrainment in conversations is likely complex and nonlinear across time, thus requiring a flexible,

---

[1]To enable clinicians to evaluate 52 conversations within a reasonable timeframe, 2 min was selected. While all SLPs agreed that 2 min was ample time to evaluate a conversation, we acknowledge that the evaluation may change over the course of the conversation as partners become familiar with one another and the dialogue task.

Figure 1. Comparison of expert clinical assessment of both healthy (reported in Borrie et al., 2019) and disordered conversations (reported herein). As evidence by the dashed line, the majority of healthy conversations were rated as more entrained (scores above 4), whereas the majority of disordered conversations were rated as less entrained (scores below 4).
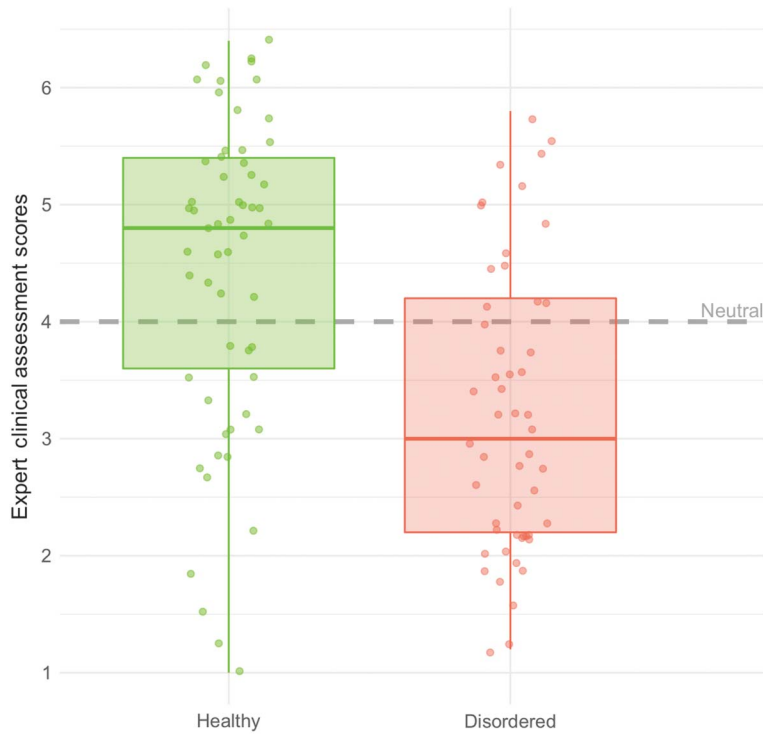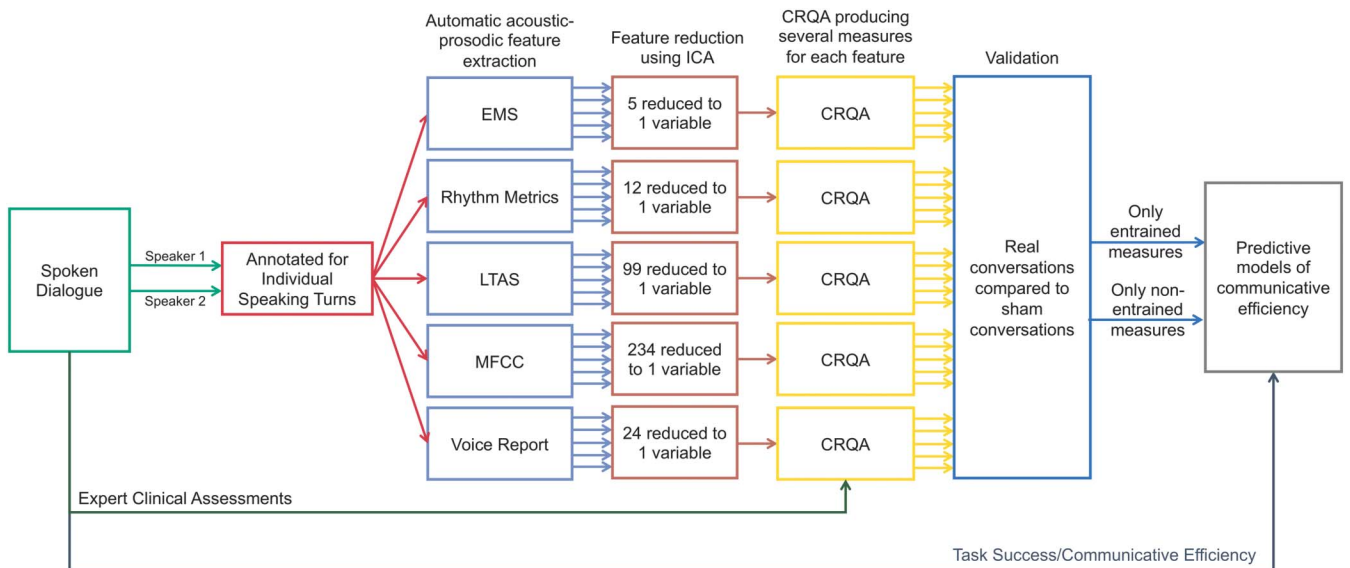

Figure 2. High-level diagram of the clinically-informed methodology for characterizing conversational entrainment in the speech domain. ICA = independent component analysis; CRQA = cross-recurrence quantification analysis; EMS = envelope modulation spectrum; LTAS = long-term average spectrum; MFCC = mel-frequency cepstral coefficient.

nonlinear analysis to adequately characterize the communication phenomenon. For several reasons, CRQA is particularly useful. First, CRQA makes no assumptions about linearity. It does this by measuring instances in time in which two time series visit similar states—this is termed *recurrence* (Coco, Dale, & Keller, 2017). That is, CRQA captures entrainment, in terms of similar repeating acoustic values between conversational participants, regardless of the time course. For example, what is happening with Participant A in the first utterance may be entrained with what Participant B is doing at a much later segment of the conversation. Second, it analyzes the entire conversation simultaneously, thereby measuring recurrence at all possible lags (i.e., all possible time delays between the conversational partners). Importantly, this means we make no assumptions about who is "leading" the conversation. Third, the approach normalizes (standardizes) the time series to be z scores. That is, the mean is 0, and the units of the feature is now standard deviations for both speakers' data. This allows the comparison across different speakers without differences that could simply result from having a female–female dyad instead of a male–female dyad. Finally, CRQA produces several interpretable measures that quantify not only the amount of entrainment but also the stability and complexity of the aligned behavior (see Fusaroli, Konvalinka, & Wallot, 2014, for a review of CRQA application in social interaction). We include four of the five[2] output measures used (and defined) in our work with healthy populations (Borrie et al., 2019) and implement the analysis using the "CRQA" package in the R statistical environment (CRQA package Version 1.0.6 and R Version 3.5.1; Coco & Dale, 2014; R Core Team, 2018). Note that the CRQA measures assessed the conversation at the utterance level, ignoring all pauses in speech.

1. *Recurrence rate* is defined as the number of single instances of alignment between the dyads accounting for the number of turns taken over the entire conversation. Higher recurrence rate values indicate higher amounts of single-instance entrainment.

2. *Sustained recurrence* is defined as the amount of alignment between the dyads that is maintained for longer than a single instance. Higher sustained recurrence values indicate higher amounts of sustained entrainment. This is also known as "nrline" in the "CRQA" package, with the following technical definition: "The total number of lines in the recurrent plot."

3. *Length* is defined as the average length/time that the dyads stay aligned with one another. Higher length values indicate entrainment for longer stretches, on average.

4. *Max length* is defined as the longest length/uninterrupted period of time that the dyads stay aligned with one another. Higher max length values indicate longer periods of entrainment.

Parameter Settings: Using Expert Clinical Assessment to Inform CRQA

For CRQA to perform measurements of entrainment, a number of parameters must be selected. These are used to define such aspects as how close the two time series must align to be counted as recurrence. Two of these parameters (delay and radius) cannot be adequately selected without prior knowledge regarding the temporal scale at which entrainment occurs or how similar acoustic features between two individuals should be to be considered entrained. The delay parameter refers to the lag of the alignment. Radius, on the other hand, refers to the threshold at which the two states are considered to be sufficiently aligned. As described in Borrie et al. (2019), we used a data-driven approach, using expert clinical assessment of entrainment, and cross-validated (10-fold) *k*-nearest neighbor models, to inform model parameters across a wide range of possible values, informed by the algorithm used by Duran and Fusaroli (2017). To do this, we used the mean expert clinical assessment scores and a binary classification of entrained (scores above 4) and not entrained (scores below 4). This resulted in 14 (27%) conversations classified as entrained and 37 (71%) conversations classified as not entrained; a single conversation (2%) was classified as neutral. Then, for each separate feature set (e.g., MFCC, EMS), we assessed several values of the delay and radius parameters. The set of parameters that resulted in the highest predictive accuracy of the cross-validated expert clinical assessments were the final parameters selected for that feature set. Thus, this approach finds the parameters that optimize the relationship between the CRQA measures and the expert clinical assessments of entrainment. The resulting measures of this approach, then, have information from the acoustic speech features and clinical assessments of the conversations. Table 1 shows the delay–radius combinations that resulted in the highest cross-validated accuracies for the conversation assessments. Notably, the CRQA embedding parameter was set to 1 in each analysis.

Using these high-accuracy parameters, the CRQA output measures were calculated for each conversation and feature set.[3] It is important to note that the clinical

---

[2]The output measure not used is known as *entropy*. It often had problems converging and thus contains many missing values. This is likely due to the parameters being set across all conversations and not individually for each conversation. Although beneficial across conversations, this may make the parameters a poorer fit for some conversations. In those cases, the more complex entropy measure may not be able to converge. Furthermore, the entropy measure was not found to be important in earlier work regarding entrainment (Borrie et al., 2019).

[3]The class asymmetry (i.e., the predicted outcome for the models has a prior probability of being unsuccessful at 71%) as found in this study can inflate the accuracy rates (chance error rate is 71% instead of 50%). However, all models for the parameter setting had information above and beyond that expected by chance (see Table 1). No weighting or resampling, based on prior probabilities, was used.

Table 1. Parameters (delay and radius) selected for each feature set based on the highest cross-validated prediction of expert clinical assessment of conversation.

Feature	Accuracy (%)	Delay	Radius
MFCC	82.2	12	3
LTAS	79.5	11	3
Voice report	79.7	8	6
EMS	76.3	5	1
Rhythm metrics	77.8	8	4

Note. MFCC = mel-frequency cepstral coefficient; LTAS = long-term average spectrum; EMS = envelope modulation spectrum.

assessment scores are simply used to identify the two optimal hyperparameters of the model (delay and radius). That is, we do not use the clinical assessment scores in any other way within CRQA.

Acoustic Entrainment Validation

To validate the objective measures of acoustic entrainment, we generated a comparison corpus of 500 sham conversations,[4] consisting of conversations between two participants, one healthy partner and one partner with dysarthria, who did not converse with one another in the original (real) corpus. These shams were created by assigning multiple, random, not-in-conversation healthy partners to each individual with dysarthria, resulting in not-in-conversation dyads (sham conversations). Thus, the sham conversations had all of the interdependent behavior of entrainment removed from the conversation, leading to measures that represent a null (or no relationship) distribution, while maintaining the other aspects of the conversation (e.g., acoustic speech signals of two individuals). By comparing the real conversations with the sham conversations for each CRQA output measure (using the same parameter settings as the real conversations), we can assess whether that measure adequately captures the coordinated behavior that presumably occurs during real conversation. This use of sham conversations to validate measures of conversational entrainment has been used previously (e.g., Bernieri et al., 1988; Duran & Fusaroli, 2017). The statistical comparisons between real and sham conversations were done using the "furniture" R package (Version 1.9.0; Barrett & Brignone, 2017), with data cleaning done using "dplyr" and visuals done using the "ggplot2" package (Wickham, 2016; Wickham, François, Henry, & Müller, 2019). Because 20 *t* tests are performed—testing the differences between real and sham conversations across the 20 feature-measure combinations—we used the Bonferroni adjustment of .0025.

---

[4]Rather than 52 sham conversations (to match the number of real conversations), 500 were used to increase certainty in our estimates and comparisons. Having a higher number of shams reduces the Type II error rates (false negatives) without any inflation of Type I error rates (false positives).

This approach keeps the Type I error rate (false positives) across all the comparisons together at 0.05. Notably, all data, code, output, and other supplemental materials from this study are available at https://osf.io/ktg5q.

Predictive Models

We then examined whether the objective acoustic measures classified as entrained (i.e., measures significantly higher in real vs. sham conversations) were predictive of an objective measure of conversational success, communicative efficiency. As done in Borrie et al. (2019), we used elastic net regression, support vector machines, and *k*-nearest neighbors for this prediction analysis. These specific approaches were chosen as they represent a breadth of linear and nonlinear approaches to prediction, they often have high prediction accuracy in various situations, and obtaining variable importance measures is straightforward. Selection of model-specific parameters of the statistical approaches was based on 10-fold cross-validation, wherein many combinations of the approach-specific parameters were tested (the default parameters to vary in the "caret" R package; Kuhn, 2019). For example, in elastic net, the alpha and lambda values are varied, with the combination that produces the highest cross-validated predictive accuracy being chosen.

Secondary Analysis of Relationships

For additional exploration of the data, we used Pearson correlations to assess potential relationships of interest with and between expert clinical assessments of entrainment, communicative efficiency, and dysarthria severity ratings, with the caveat that the data collection for this study was not set up to target severity (i.e., unequal spread of participants with mild, mild-to-moderate, moderate, and severe dysarthria). Finally, we used a chi-square analysis to ensure that expert clinical assessment of entrainment was not influenced by the gender composition of the conversational dyad (i.e., female–female, male–female, male–male).

Results
Objective Acoustic Evidence of Entrainment

The CRQA produced four output measures for each acoustic feature set, operationally termed *feature measures*. A summary of the feature measures for both real and sham conversations can be found in Table 2. The table reports the *p* values from independent-samples *t* tests—adjusting for any instances of unequal variances along with their corresponding standardized mean difference effect sizes. The table shows objective evidence of acoustic entrainment in only the single phonatory dimension of speech signal behavior, voice report–sustained recurrence ($p < .001$). All other feature measures were nonsignificant at the conservative Bonferroni-adjusted level

of .0025.[5] Thus, the goal-oriented conversations between participants with dysarthria and healthy participants (disordered conversations) were characterized by less acoustic entrainment than goal-oriented conversations between two healthy participants (healthy conversations; Borrie et al., 2019), as illustrated in Figure 3.

Predictive Models

Across the 52 goal-oriented conversations collected, communicative efficiency scores ranged from 4 to 22 ($M = 12.3$, $SD = 4.3$). Using the single feature measure that entrained (voice report–sustained recurrence), we fit three statistical models, assessing their ability to predict communicative efficiency scores. Using the R^2 of each model, voice report–sustained recurrence explained between 28% and 31% of the variance of communicative efficiency. Notably, the R^2 value is like that used in linear regression, which provides us with the amount of variance explained by the model. This can be conceptualized as the correlation (squared) of the predicted values and the actual observed values. That is, higher R^2 means the predicted values are consistently closer to the observed values than for lower values of R^2.

Secondary Analysis of Relationships

The first relationship of interest, between expert clinical assessments of entrainment and communicative efficiency, was significant, $r(52) = .422$, $p = .002$, corroborating the observed predications between acoustic entrainment and communicative efficiency. The next relationships of interest, between dysarthria severity and all other measures, were not strong: acoustic measures of entrainment, $r(52) = .191$, $p = .175$; clinical assessment of entrainment, $r(52) = -.002$, $p = .991$; and communicative efficiency, $r(52) = -.002$, $p = .991$. Finally, clinical assessment of entrainment was not related to the gender composition of the conversational partners dyad, $\chi^2(2) = 0.93$, $p = .630$.

---

[5]Two additional analyses were done to further corroborate the methodology. Firstly, we performed the CRQA using just the first 2 min of the conversations (the data rated by the clinicians) and found that the results were virtually identical—no acoustic entrainment in articulation and rhythmic dimensions of speech. The small difference observed is that, while we found phonatory entrainment when all data were analyzed (i.e., 10 min), we did not find this entrainment in analysis of the first 2 min. This is likely due to limited data available for the analysis and the small effect size associated with entrainment along the phonatory dimension. Thus, the results are not substantially affected by analyzing 2 min versus 10 min of conversation. Secondly, the CRQA results were also tested using continuous clinical assessment (instead of the binary—entrained or not entrained—measure reported herein). The results showed similar levels of entrainment in the corpus and had similar predictive accuracy of conversational efficiency. The binary measure is reported given it is the methodology employed in the first study with healthy conversations.

Table 2. Comparison of real conversations with sham conversations by feature set.

Variable	Real n = 52	Sham n = 500	p value
Articulation			
MFCC			
Length	1.980 (0.283)	1.965 (0.342)	.765
Max length	2.173 (0.550)	2.074 (0.499)	.178
Sustained recurrence	9.558 (9.348)	6.244 (4.333)	.014
Recurrence rate	1.859 (0.306)	1.869 (0.205)	.826
LTAS			
Length	1.998 (0.326)	1.977 (0.302)	.630
Max length	2.250 (0.682)	2.158 (0.534)	.251
Sustained recurrence	29.654 (61.085)	11.094 (18.270)	.034
Recurrence rate	2.268 (1.154)	2.257 (0.867)	.932
Phonation			
Voice report			
Length	2.063 (0.113)	2.046 (0.052)	.286
Max length	2.673 (0.648)	2.684 (0.607)	.902
Sustained recurrence	57.288 (47.083)	29.828 (23.718)	< .001
Recurrence rate	4.225 (0.943)	4.033 (0.891)	.140
Rhythm			
EMS			
Length	1.308 (0.961)	1.096 (1.002)	.137
Max length	1.308 (0.961)	1.100 (1.008)	.145
Sustained recurrence	1.077 (1.064)	0.890 (1.090)	.239
Recurrence rate	0.629 (0.193)	0.688 (0.081)	.035
Rhythm metrics			
Length	1.995 (0.303)	2.011 (0.189)	.596
Max length	2.250 (0.556)	2.280 (0.553)	.710
Sustained recurrence	30.250 (60.078)	16.864 (35.464)	.017
Recurrence rate	2.647 (1.134)	2.864 (1.238)	.227

Note. MFCC = mel-frequency cepstral coefficient; LTAS = long-term average spectrum; EMS = envelope modulation spectrum.

Discussion

Here, a multidimensional methodology, developed and validated in an earlier study with healthy populations, is used to characterize the communication phenomenon of conversational entrainment in a large corpus of goal-oriented conversations between individuals with dysarthria and healthy participants. We observed evidence of acoustic entrainment in phonatory, but not rhythmic and articulatory, behavior. This differs from conversations between healthy participants, which showed acoustic entrainment across phonatory, rhythmic, and articulatory dimensions of speech (Borrie et al., 2019). As discussed, in order to entrain, individuals must be able to produce, perceive, and modify their speech production behavior. Dysarthria is characterized by speech production deficits in rhythmic (e.g., reduced stress), articulatory (e.g., imprecise articulation), and phonatory behavior (e.g., monotone), causing us to postulate that its very presence would disrupt conversational entrainment. The hypothesis of entrainment impairments in dysarthria is further advanced by preliminary evidence

Figure 3. Comparison of acoustic entrainment (sustained recurrence) in both healthy (reported in Borrie et al., 2019) and disordered conversations (reported herein) in terms of the standardized effect sizes. Effect sizes denoted with an asterisk are those that were found to be significant in each study. MFCC = mel-frequency cepstral coefficient; LTAS = long-term average spectrum; EMS = envelope modulation spectrum.
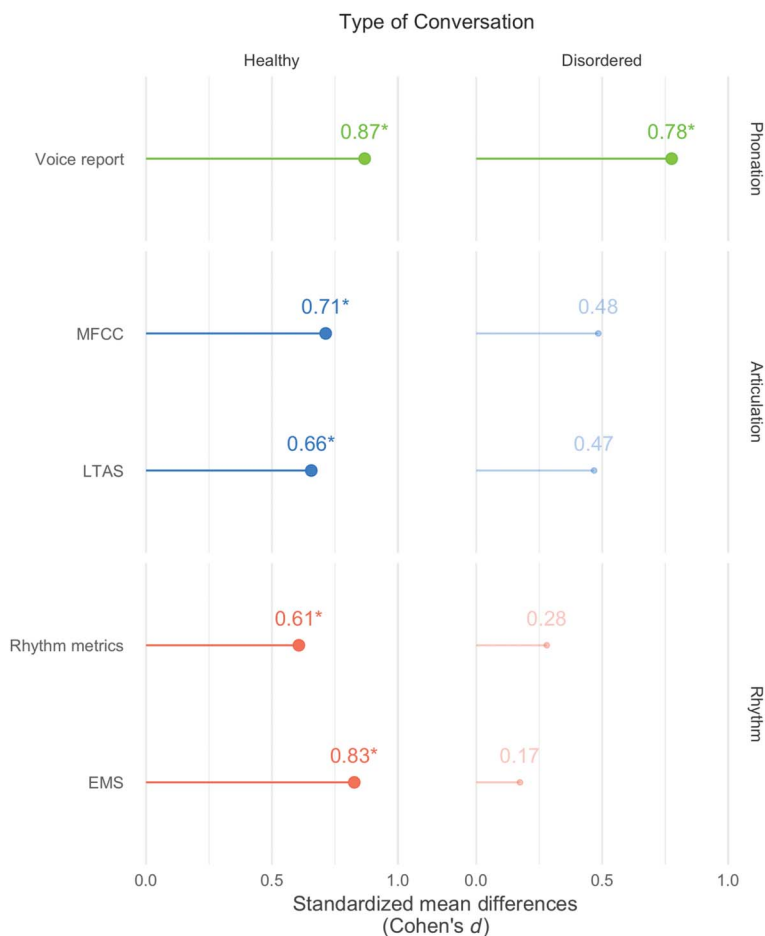
of deficits with basic acoustic speech features and simple turn-by-turn measures of local entrainment (Borrie & Liss, 2014; Borrie et al., 2015) and current expert clinical assessment of the conversations, where the majority fared poorly in terms of holistic impressions of entrainment. Thus, this study affords additional evidence that, in general, the pathological production parameters of dysarthric speech disrupts conversational entrainment.

We do, however, observe entrainment of phonatory behavior, a speech signal dimension largely based on pitch properties. We see two possible explanations for this result. The first is that individuals with dysarthria modify their phonatory behavior to more closely resemble that of their healthy partners' productions. Certainly, the production deficits of dysarthria do not preclude deficits in modifying speech behavior; indeed, the notion that individuals with dysarthria can shift their speech in the direction of more healthy/typical productions is the rational for speaker-oriented interventions in the clinic (e.g., contrastive stress drills, loud speech; see Duffy, 2015). The second explanation,

not mutually exclusive from the first, is that the healthy interlocuter modifies their phonatory behavior to more closely resemble that of the dysarthric productions. This explanation is supported by research conducted in highly controlled, experimental paradigms, showing that healthy individuals can modify their pitch properties to align more closely with the pathological speech parameters of dysarthria, both with (Borrie & Schäfer, 2015) and without explicit instruction (Borrie & Liss, 2014). While directionality of entrainment (i.e., who is modulating their speech—the individual with dysarthria and/or the healthy interlocuter) cannot be realized within the scope of the current analysis—CRQA, in the way we performed it, simply determines when two states are similar—the current findings indicate that, in ecologically valid interactional settings and in the absence of any overt directive, phonatory behavior is, overall, amendable to production shifts, arising from the individual with dysarthria and/or the healthy interlocuter. We speculate that phonatory behaviors may be the easiest speech dimension to modify due to the overt nature of

pitch manipulations; however, future studies are required to substantiate this claim.

As per our earlier work investigating conversations between two healthy participants, we found that acoustic entrainment, which in the current study was solely phonatory entrainment, was predictive of an objective measure of conversational success, communicative efficiency. This predictive relationship was corroborated by a positive relationship between expert clinical assessments of entrainment and communicative efficiency. The link between acoustic entrainment and communicative efficiency in the context of dysarthria is also consistent with the results of an earlier study, where entrainment is captured as a local turn-by-turn correlation measure of pitch synchrony (Borrie et al., 2015). Here, we afford confirmatory evidence of this relationship with clinically informed measures of entrainment and a much more global characterization of speech synchronization. Thus, entrainment appears to play a role in conversational success in dysarthria.

It is important to acknowledge that our gross measure of communicative efficiency, operationally defined as task success in the goal-directed conversations, may not be the optimal metric for rating success in conversations involving individuals with dysarthria. A key characteristic of dysarthria is speaking rate abnormalities, with many speakers presenting with slow speech rate (Weismer & Kim, 2010). As task success here is quantified by the number of differences that are identified in 10 min of spoken dialogue, speaking rate could influence communicative efficiency scores. That is, individuals with dysarthria and slower speaking rates may be differentially disadvantaged, regardless of whether they achieve some level of phonatory entrainment with their healthy conversational partner. This points to the idea that aspects considered to contribute to successful conversation in healthy populations may be less applicable to populations with communication disorders, particularly when rate is disturbed. Future work in this area of study will include more comprehensive and inclusive measures of conversational success.

We did not observe evidence of a strong relationship between dysarthria severity and acoustic entrainment, wherein conversations involving mild dysarthria showed similar levels of phonatory entrainment as conversations involving moderate or severe dysarthria. This pattern, where severity and entrainment measures showed no significant association, was also evident in the expert clinical assessment scores. Furthermore, severity and communicative efficiency showed no significant association. At first blush, this pattern of findings is counterintuitive, as it might be expected that conversations involving mild dysarthria would have less impact on entrainment and conversational success than those involving severe dysarthria. There are at least two potential explanations of this finding. First is that this counterintuitive finding is the result of the study design, wherein we did not control for dysarthria severity and the distribution was heavily skewed toward the mild end of the continuum. Thus, a study appropriately controlling for severity may indeed yield the expected relationships. A second possible explanation is that the global impression of "dysarthria severity" does not capture differences in conversations most relevant to entrainment. For example, one could imagine relatively well-preserved entrainment with a speaker who is regarded as severely dysarthric because of largely consistent speech degradation such as slow rate, monotone, strained–strangled phonation, and articulatory imprecision. Conversely, a speaker with mild dysarthria may present with largely inconsistent and thus unpredictable speech timing disturbances that greatly interfere with entrainment. Indirect support for a role of signal predictability in entrainment comes from studies investigating perceptual learning of dysarthric speech, whereby acoustic regularity is key for listener adaptation to the degraded signal (Borrie, Lansford, & Barrett, 2018; Lansford, Borrie, & Barrett, 2019). We postulate that the relationships among dysarthria presentation (i.e., severity, acoustic regularity, deviant speech characteristics), conversational entrainment, and communicative success are likely highly complex and recognize the need for investigations that methodically investigate these associations.

Limitations and Future Directions

It is important to acknowledge some of the limitations of the methodology employed. First, our approach to using CRQA to quantify global acoustic entrainment in a conversation does not use any time-alignment estimation, a method that others have employed when using CRQA to analyze entrainment of speech rate (Duran & Fusaroli, 2017). While this was done specifically to avoid an additional estimation procedure, it limits our ability to assess directionality of entrainment (i.e., who is modulating their speech). CRQA, in the way we performed it, simply determines when two states are similar. Furthermore, the current methodology takes a binary perspective to the concept of conversational entrainment—it is either present or absent. Yet, expert clinical assessment, where scores spanned the full 1–7 scale, suggests that entrainment is a continuous phenomenon where some conversations are more/less in sync than others. In addition, with expert clinical assessment scores above 4 for 14 of the 52 conversations, approximately 30% of the spoken dialogue samples were in fact evaluated as having an appropriate level of entrainment in terms of cohesiveness, connection, and conversational flow. So, while the current methodology yields big picture ideas regarding entrainment in the context of dysarthria—overall, less entrainment relative to healthy conversations and a relationship with conversational success—it has limited value for investigating the nuances and idiosyncrasies of behavioral synchrony in conversation. For example, beyond directionality, what makes conversational entrainment possible in some healthy–dysarthric dyads? Do patterns of entrainment change and develop over time, as participants become familiar with one another? What degree of entrainment is actually optimal for conversational success? Absolute behavioral mimicry is, of course, not the goal. Furthermore, Fusaroli, Raczaszek-Leonardi,

and Tylén (2014) have advanced a theoretical framework of *dialog as interpersonal synergy*, in which behavioral synchrony consists of both aligned (same/similar) and complementary (different but still interdependent) behaviors, modulated in context-sensitive ways. Indeed, a functional role for complementary behavior may be particularly applicable when exploring conversational entrainment in communication disorders, where optimally aligned behavior may be difficult to achieve. In such cases, are there compensatory behaviors (e,g., precise articulation) that healthy interlocuters engage in (or be trained to engage in) to enhance conversational success? These sorts of questions may or may not be overly important for investigations with healthy populations but will be imperative for understanding entrainment in clinical populations, particularly as we consider potential management of deficits. Thus, we advance that the development and refinement of methodologies for characterizing entrainment in communication disorders must continue, with a concerted focus on clinical utility.

Conclusion

Using a methodology that weaves together automatic acoustic feature extraction, feature reduction, recurrence quantification, and expert clinical assessment, we characterize acoustic entrainment in a large corpus of experimentally elicited conversations between individuals with dysarthria and healthy interlocuters. Overall, we find less acoustic evidence of entrainment in these conversations relative to that evident in historical data from the same methodological application in conversations involving two healthy interlocuters. Entrained behavior did, however, predict communicative efficiency of the conversations, evidencing a functional role in conversational outcomes. Thus, the current findings advance entrainment deficits as an important variable in dysarthria that may have causative effects on the success of everyday communication and, consequently, quality of life.

Acknowledgments

References

Bailenson, J. N., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science, 16,* 814–819.

Barrett, T. S., & Brignone, E. (2017). Furniture for quantitative scientists. *The R Journal, 9,* 142–148.

Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation, 6*(4), 802–813.

Berisha, V., Liss, J., Sandoval, S., Utianski, R., & Spanias, A. (2014). Modeling pathological speech perception from data with similarity labels. *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference, 2014,* 915–919.

Bernieri, F., Reznick, J., & Rosenthal, R. (1988). Synchrony, pseudo synchrony, and dissynchrony: Measuring the entrainment process in mother–infant interactions. *Journal of Personality and Social Psychology, 54,* 243–253.

Bloch, S., & Wilkinson, R. (2009). Acquired dysarthria in conversation: Identifying sources of understandability problems. *International Journal of Language & Communication Disorders, 44,* 769–783.

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [Computer program]. Version 6.0.39. Retrieved from http://www.praat.org/

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,* 283–296.

Borrie, S. A., Lansford, K. L., & Barrett, T. S. (2018). Understanding dysrhythmic speech: When rhythm does not matter and learning does not happen. *The Journal of the Acoustical Society of America, 143,* EL379–EL385.

Borrie, S. A., & Liss, J. M. (2014). Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research, 57,* 815–824.

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.

Borrie, S. A., & Schäfer, M. C. M. (2015). The role of somatosensory information in speech perception: Imitation improves recognition of disordered speech. *Journal of Speech, Language, and Hearing Research, 58,* 1708–1716.

Cleveland, T. F., Sundberg, J., & Stone, R. E. (2001). Long-term-average spectrum characteristics of country singers during speaking and singing. *Journal of Voice, 15,* 54–60.

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

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology, 5,* 510.

Coco, M. I., Dale, R., & Keller, F. (2017). Performance in a collaborative search task: The role of feedback and alignment. *Topics in Cognitive Science, 10,* 55–79.

Comon, P. (1992). Independent components analysis. In J. L. Lacoume (Ed.), *Higher order statistics* (pp. 29–38). Oxford, United Kingdom: Elsevier. Retrieved from https://hal.archives-ouvertes.fr/hal-00346684

Comrie, P., Mackenzie, C., & McCall, J. (2001). The influence of acquired dysarthria on conversational turn-taking. *Clinical Linguistics & Phonetics, 15,* 383–398.

Dellwo, V., & Fourcin, A. (2013). Rhythmic characteristics of voice between and within languages. *Revue Tranel (Travaux-neuchâtelois de linguistique), 59,* 87–107.

Duffy, J. R. (2015). *Motor speech disorders: Substrates, differential diagnosis, and management* (3rd ed.). St. Louis, MO: Elsevier.

Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PLOS ONE, 2*(6), e0178140.

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,* 1043–1054.

Fusaroli, R., Konvalinka, I., & Wallot, S. (2014). The promises and challenges of using cross recurrence quantification analysis. In N. Marwan, M. Riley, A. Giuliani, & C. Webber Jr. (Eds.), *Translational recurrences. Springer proceedings in mathematics & statistics* (Vol. 103). New York, NY: Springer.

Fusaroli, R., Raczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology, 32,* 147–157.

Gordon, R. G., Rigon, A., & Duff, M. C. (2015). Conversational synchrony in the communicative interactions of individuals with traumatic brain injury. *Brain Injury, 29*(11), 1300–1308.

Kuhn, M. (2019). Caret: Classification and regression training. *R Package Version, 6,* 0–84. https://CRAN.R-project.org/package=caret

Lansford, K. L., Borrie, S. A., & Barrett, T. S. (2019). Regularity matters: Unpredictable speech degradation inhibits adaptation to dysarthric speech. *Journal of Speech, Language, and Hearing Research, 62*(12), 4282–4290. https://doi.org/10.1044/2019_JSLHR-19-00055

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

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In R. Pieraccini & A. Colombo (Eds.), *Proceedings of interspeech*. Brisbane, Australia: International Speech Communications Association.

Liss, J. M., LeGendre, S., & Lotto, A. J. (2010). Discriminating dysarthria type from envelope modulation spectra. *Journal of Speech, Language, and Hearing Research, 53,* 1246–1255.

Marchini, J. L., Heaton, C., & Ripley, B. D. (2017). *FastICA: FastICA algorithms to perform ICA and projection pursuit (R package Version 1.2-1)*. Retrieved from https://CRAN.R-project.org/package=fastICA

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. In *Proceedings of the 46th annual meeting of the Association for Computational Linguistics on Human Language Technologies: Short papers* (pp. 169–172). Stroudsburg, PA: Association for Computational Linguistics.

Palmer, A. D., Carder, P. C., White, D. L., Saunders, G., Woo, H., Graville, D. J., & Newsome, J. T. (2019). The impact of communication impairments on the social relationships of older adults: Pathways to psychological well-being. *Journal of Speech, Language, and Hearing Research, 62,* 1–21.

Phillips-Silver, J., Aktipis, C. A., & Bryant, G. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Perception, 28,* 3–14.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–190.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

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.

Tu, M., Berisha, V., & Liss, J. (2017). Interpretable objective assessment of dysarthric speech based on deep neural networks. In F. Lacerda (Ed.), *Proceedings of interspeech* (pp. 1849–1853). Baixas, France: International Speech Communication Association.

Tu, M., Jiao, Y., Berisha, V., & Liss, J. M. (2016). Models for objective evaluation of dysarthric speech from data annotated by multiple listeners. In M. B. Matthews (Ed.), *50th Asilomar Conference on Signals, Systems and Computers, 2016* (pp. 827–830). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Weismer, G., & Kim, Y.-J. (2010). Classification and taxonomy of motor speech disorders: What are the issues? In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research*. Oxford, United Kingdom: Oxford University Press.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A grammar of data manipulation. R package Version 0.8.1.* https://CRAN.R-project.org/package=dplyr

Willi, M. M., Borrie, S. A., Barrett, T. S., Tu, M., & Berisha, V. (2018). A discriminative acoustic-prosodic approach for measuring local entrainment. *Proceedings of Interspeech,* 581–585.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review, 12*(6), 957–968.

Wynn, C. J., Borrie, S. A., & Sellars, T. (2018). Speech rate entrainment in children and adults with and without autism spectrum disorder. *American Journal of Speech-Language Pathology, 27,* 965–974.

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,* 510–540.

Zbilut, J. P., Giuliani, A., & Webber, C. L., Jr. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A, 246,* 122–128.