Published in final edited form as:
J Speech Lang Hear Res. 2014 June 1; 57(3): 815–824. doi:10.1044/2014_JSLHR-S-13-0149.

Rhythm as a Coordinating Device: Entrainment With Disordered Speech

Stephanie A. Borrie[a] and **Julie M. Liss**[a]
[a]Arizona State University, Tempe

Abstract

Purpose—The rhythmic entrainment (coordination) of behavior during human interaction is a powerful phenomenon, considered essential for successful communication, supporting social and emotional connection, and facilitating sense-making and information exchange. Disruption in entrainment likely occurs in conversations involving those with speech and language impairment, but its contribution to communication disorders has not been defined. As a first step to exploring this phenomenon in clinical populations, the present investigation examined the influence of disordered speech on the speech production properties of healthy interactants.

Method—Twenty-nine neurologically healthy interactants participated in a quasi-conversational paradigm, in which they read sentences (*response*) in response to hearing prerecorded sentences (*exposure*) from speakers with dysarthria (n = 4) and healthy controls (n = 4). Recordings of read sentences prior to the task were also collected (*habitual*).

Results—Findings revealed that interactants modified their speaking rate and pitch variation to align more closely with the disordered speech. Production shifts in these rhythmic properties, however, remained significantly different from corresponding properties in dysarthric speech.

Conclusion—Entrainment offers a new avenue for exploring speech and language impairment, addressing a communication process not currently explained by existing frameworks. This article offers direction for advancing this line of inquiry.

Keywords

rhythm; communication; interaction; entrainment

Rhythmic entrainment, operationally defined as the "spatiotemporal coordination resulting from rhythmic responsiveness to a perceived rhythmic signal" (Phillips-Silver, Aktipis, & Bryant, 2010, p. 5), is a pervasive communication phenomenon, describing the alignment of behaviors during human interaction. Put more simply, conversational partners naturally adapt their verbal and non-verbal actions to more closely resemble one another. This entrainment of behavior is considered critical for successful communication and may in fact be "essential for our survival as social beings" (Gill, 2012, p. 111). We theorize that

---

Correspondence to Stephanie Anna Borrie: sborrie@asu.edu.

Disclosure: The authors have declared that no competing interests existed at the time of publication.

communication disorders induce disruptions to rhythmic entrainment in human interaction and initiate a first step to exploring this hypothesis within the context of disordered speech. Identifying breakdowns in human interaction is fundamental to ongoing efforts to delineate an evidence base for assessment and rehabilitation within the clinical environment.

Entrainment is evidenced in many aspects of spoken language, including acoustic and prosodic speech features such as speaking rate (e.g., Giles, Coupland, & Coupland, 1991; Local, 2007; Street, 1984; Webb, 1969, 1972), latency and utterance durations (e.g., Matarazzo & Wiens, 1967; Matarazzo, Weitman, Saslow, & Wiens, 1963), pitch properties (e.g., Gregory, 1990; Lee et al., 2010; Levitan et al., 2012; Local, 2007), vocal intensity (e.g., Coulston, Oviatt, & Darves, 2002; Local, 2007; Natale, 1975), voice quality (e.g., Levitan & Hirschberg, 2011; Neumann & Stack, 2000), and dialectal features (e.g., Giles, 1973). Beyond speech properties, entrainment is observed in phonological and phonetic features (e.g., Babel, 2012; Pardo, 2006), linguistic style (e.g., Danescu-Niculescu-Mizil, Gamon, & Dumais, 2011; Niederhoffer & Pennebaker, 2002), lexical choice (e.g., Brennan & Clark, 1996; Garrod & Anderson, 1987), and syntactic structure (e.g., Branigan, Pickering, & Cleland, 2000; Reitter, Moore, & Keller, 2006). Alignment of behavior extends across the motor system to include body posture, facial expression, and gesture (e.g., Furuyama, Hayashi, & Mishima, 2005; Louwerse, Dale, Bard, & Jeuniaux, 2012; Richardson, Marsh, & Schmit, 2005; Shockley, Santana, & Fowler, 2003), and is even reported in breathing (e.g., McFarland, 2001) and laughter patterns (e.g., Truong & Trouvain, 2012). Further, entrained action is supported neurologically by a growing body of evidence showing that neural oscillations in the auditory cortex are modulated in phase to match rhythmic properties of incoming speech (e.g., Lakatos et al., 2005; Peelle, Gross, & Davis, 2013).

That entrainment is such a robust and multilevel communication phenomenon gives rise to the question of its functional utility. According to cognitive theories of interaction, entrainment is central to understanding spoken language (Pickering & Garrod, 2004), supporting sense-making and information exchange (Gill, 2012). For example, coordination of behavior relies on the ability to detect rhythmicity, which enables the listener to generate temporal predictions about the acoustic and phonetic realizations of upcoming speech (e.g., Brown, Salverda, Dilley, & Tanenhaus, 2011). *Predictions*, or expectations of how the conversation will unfold, reduce the cognitive effort associated with speech processing. There is also evidence that entrainment aids memory for characteristics of one's communication partner (Macrae, Duffy, Miles, & Lawrence, 2008; Woolhouse & Tidhar, 2010), which has been shown to improve perceptual processing of speech (Goldinger, 1998; Mattys & Liss, 2008). Entrainment is also important to fulfill a number of social and emotional functions of human interaction. It binds turn-taking dynamics and is positively associated with increased overlaps, reduced interruptions, and shorter interturn latencies (e.g., Local, 2007; Wilson & Wilson, 2005). Further, the alignment of behavior during interaction is considered fundamental to the development of rapport, empathy, and intimacy between conversational partners (e.g., Bailenson & Yee, 2005; Chartrand & Bargh, 1999; Lee et al., 2010; Miles, Nind, & Macrae, 2009; Putman & Street, 1984; Smith, 2008; Street & Giles, 1982). Entrainment, therefore, serves as a powerful coordinating device, uniting individuals in time and space to optimize comprehension, establish social presence, and create positive and satisfying relationships.

Entrainment is clearly critical to successful communication, and although this interaction phenomenon has, and continues to be, studied widely across a number of disciplines including psychology, linguistics, sociology, evolutionary biology, neurology, and musicology, entrainment of spoken language has received very little attention in the field of speech-language pathology and the study of communication disorders (see Buder & Edrington, 2008, for a case study with aphasia and Nakano, Kato, & Kitazawa, 2011, regarding lack of eye-blink entrainment in autism spectrum disorders). We, however, consider entrainment as fundamental to the study of communication breakdowns in clinical populations. According to the sensorimotor theory of “beat induction” proposed by Todd, Lee, and O'Boyle (2002) and extended to the social domain by Phillips-Silver, Aktipis, and Bryant (2010), the capacity to entrain is dependent on three critical components: (a) rhythmic detection; (b) rhythmic action; and (c) rhythmic integration. Entrainment, therefore, necessitates the ability to perceive rhythmic information, to produce rhythmic information, and to integrate rhythmic information—that is, to adjust one's motor output in response to sensory input. In theory, a breakdown in any or all elements of this model will disrupt entrainment and negatively influence human communication. Given that rhythmic deficits are widespread in a number of communication disorders, it is likely that disruptions in entrainment are a common feature.

In the present report, we focus on *dysarthria*, a speech production disorder that manifests itself in a number of perceptual and acoustic disturbances that compromise the integrity of the acoustic signal. Key is abnormal speech rhythm, evident in a number of phonatory and articulatory processes, including those related to the production of pitch (e.g., pitch breaks, monotone), stress (e.g., reduced stress, excess and equal stress), loudness (e.g., reduced loudness, monoloudness), and speaking rate (e.g., short phrases, variable rate). These speech degradations contribute to reductions in speech intelligibility, meaningful communication, and quality of life (Yorkston, 1996, 2007). While we have, as a discipline, addressed the important roles of speech production and listener needs in assessment and intervention of intelligibility deficits, we have not experimentally established how the rhythmic disturbances that characterize disordered speech may interfere with entrainment and impact human communication. The most basic consequence of impaired entrainment is that the comprehension benefit afforded by aligning rhythmic behavior may not be realized. In this sense, deficits or disruptions in entrainment may contribute to the intelligibility decrement of the disorder by drawing on (rather than optimizing) cognitive resources, increasing the effort required to decipher the degraded input. Another outcome is an undesirable influence on social presence and interpersonal relations. Accordingly, entrainment deficits in clinical populations may reveal an important, yet unexplored, source of communication difficulty and demise in quality of life.

Current Study

The current study represents a first step in evaluating the contribution of disordered speech to rhythmic entrainment by asking the question: Does dysarthric speech induce production changes, manifested in speech rate and pitch variation,[1] in the speech of healthy interactants?[2] On the basis of the rich literature that underscores the pervasiveness of entrainment in typical human interaction, it was hypothesized that dysarthric speech would induce, at least to some degree, a rhythmic shift in the spoken productions of healthy people. Whether healthy interactants modify their speech to more closely align with the abnormal rhythmic patterns that characterize disordered speech offers valuable insight into entrainment in disordered settings and supports examination of this communication phenomenon as a viable target of investigation in the domain of speech-language pathology.

Method

Design

A within- and between-subjects design was used to investigate production adjustments in rhythmic properties (speech rate and pitch variation) of speech produced by healthy individuals (interactants) exposed to neurologically degraded (dysarthric) and neurologically healthy (control) speech stimuli, in a quasi-conversational exposure-response paradigm.

Participants

Data were collected from 29 healthy women ages 19–39 years (*M* = 24.41; *SD* = 5.01). All participants were native speakers of American English and reported no significant history of contact with persons having motor speech disorders and no identified language, learning, hearing, or cognitive disabilities. Participants were recruited from undergraduate classes at Arizona State University and received course credit for their participation in the study.

Speech Stimuli

Speech material consisted of dysarthric and control sentence productions, derived from a larger stimuli set and described in detail elsewhere (Liss et al., 2009). Sentences elicited from four female speakers with dysarthria, including hypokinetic dysarthria associated with Parkinson's disease (*n* = 2) and ataxic dysarthria associated with cerebellar disease (*n* = 2), as well as four neurologically healthy ageand gender-matched controls, were used. All speakers with dysarthria exhibited a speech disorder of moderate severity and were selected based on the presence of cardinal speech features associated with the corresponding speech diagnosis.[3] The speakers ranged from age 45 years to 60 years. The reader is directed to Liss et al. (2009) for further details on the speakers and a comprehensive description of the stimuli collection protocol.

The stimulus set for the current study consisted of five full sentences produced by each of the eight speakers included in the study, yielding a total of 40 sentence productions. These

---

[1] Speech rate and pitch variation were selected for two reasons: (a) they contribute to the perceptual experience of rhythm (Kohler, 2009) and (b) abnormalities are a common feature of the dysarthrias.

[2] The term *interactant* refers to the experimental design, in which participants read aloud sentences on a screen after first hearing a different sentence produced by another person. Although there is no instruction to interact or even to attend to the other sentences, the turn-taking nature of the task is interactive.

[3] Cardinal deviant speech features exhibited by speakers with hypokinetic dysarthria included rapid articulation rate, rushes of speech, imprecise articulation, monopitch, and reduced loudness. Cardinal deviant speech features exhibited by speakers with ataxic dysarthria included “scanning” speech, imprecise articulation with irregular articulatory breakdown, irregular monopitch, and loudness changes. See Liss et al. (2009) for additional details.

sentences (exposure productions), composed of 9–12 words (15–17 syllables), were adapted by White and Mattys (2007) from a larger set of stimuli developed specifically for studies on speech rhythm (Nazzi, Bertoncini, & Mehler, 1998). The five sentences are reported in Appendix A.

Using Praat (Boersma & Weenink, 2013) and standard operational definitions and procedures (Peterson & Lehiste, 1960; Weismer, 1984), all 40 exposure productions were analyzed by the first author for a measure of duration and fundamental frequency variation. Duration of each phrase was then divided by the number of syllables in the phrase to calculate speech rate, reported in syllables per second (sps). Twenty-five percent of the exposure productions were remeasured by the first author (intrajudge) and by a second trained judge (interjudge) to obtain reliability estimates for the acoustic metrics. Discrepancies between the remeasured data and the original data revealed that agreement was high (all correlations $r > .95$), with only minor absolute differences. Figures 1 and 2 show box and whisker plots of speech rate and pitch variation, respectively, of the exposure productions associated with hypokinetic, ataxic, and control speech.

Figure 1 reflects the speech rate for exposure productions by speech type. A one-way analysis of variance (ANOVA) revealed a significant effect of speech type on speech rate, $F(2, 37) = 168.45$, $p < .001$, $\eta^2 = .90$. Post hoc tests, using Bonferroni correction, revealed that all comparisons were significant. The speech rate of the ataxic productions was significantly slower than both the control, $t(28) = 15.57$, $p < .001$, Cohen's $d = 1.97$, and the hypokinetic productions, $t(28) = 15.33$, $p < .001$, $d = 1.88$, and the speech rate of the hypokinetic productions was significantly faster than the control productions, $t(28) = 15.57$, $p < .001$, $d = 1.96$. Figure 2 reflects the average pitch variation for exposure productions by speech type. A one-way ANOVA showed a significant effect of speech type on pitch variation, $F(2, 37) = 99.18$, $p < .001$, $\eta^2 = .84$. Post hoc tests, using Bonferroni correction, revealed that the pitch variation of the control stimuli was significantly greater than that of both the hypokinetic, $t(28) = 9.93$, $p < .001$, $d = 1.84$, and ataxic stimuli, $t(28) = 10.80$, $p < .001$, $d = 1.87$. There was no significant difference in pitch variation between the disordered speech types. Thus, these metrics validate that the hypokinetic productions were associated with a fast speech rate and reduced pitch variation and that the ataxic productions were associated with a slow speech rate and reduced pitch variation. The data are used for comparative purposes in the Results section. The same five sentences used as stimuli for the exposure productions, described in Appendix A, were used to elicit speech from the interactants, as outlined in the experimental procedure below.

Experimental Procedure

The experiment was conducted in the Motor Speech Disorders Laboratory at Arizona State University. Interactants were all blinded to the purpose of the study. Upon entering the laboratory, interactants were seated in a chair and fitted with sound-attenuating headphones (Sennheiser AD 280 pro) and a head-mounted microphone, positioned at a 5-cm mouth-to-microphone distance. The experiment was presented via a laptop computer, preloaded with the experimental procedure and recording software. Interactants were informed that task-specific instructions would be provided in the computer presentation.

The experimental procedure consisted of two tasks. The first task required interactants to simply produce speech. Interactants were presented with a slide with the instructions to "please read aloud the sentences listed on the following slides." Each of the five sentences was presented twice, so that each interactant produced a total of 10 sentences (habitual productions). The second task involved an exposure-response paradigm in which interactants were presented with an exposure production produced by one of the eight speakers, via the headphones, and followed by a slide on the screen prompting them to read aloud a different sentence on the screen (response productions). Interactants were not informed of the purpose of this experiment, nor were they explicitly instructed to attend to the exposure productions of the dysarthric or control speech. This exposure-response continued in a cyclic fashion until interactants had been presented all 40 of the exposure productions (5 sentences × 8 speakers) and, themselves, produced a total of 40 response productions. Presentation of the exposure productions was blocked by a speaker, so that interactants heard all five productions from one speaker before proceeding to the next speaker. Speaker order was randomized across interactants. Within each speaker, the presentation order of the exposure productions was also randomized, as was the order of the response productions elicited by the interactant. It is important to emphasize that exposure and response sentences alternated in a quasi-random fashion such that the response production was never the same as the exposure production (i.e., no mimicry condition). Habitual productions and response productions elicited during the experimental procedure were recorded digitally to a laptop computer using Sony Sound Forge (Media Software, Madison, WI) at 48 kHz (16-bit sampling rate). On completion of the experiment, interactants were debriefed with a short description regarding the nature of the study.

Acoustic Analysis

The total data set from the experimental procedures consisted of 50 sentence productions (10 habitual productions and 40 response productions) for each of the 29 interactants, yielding a total of 1,450 sentence productions. Using identical procedures and definitions to those employed in acoustic analysis of the speech stimuli, each individual sentence production was analyzed for a measure of speech rate and pitch variation using Praat (Boersma & Weenink, 2013).

Reliability of Acoustic Analysis

Twenty-five percent of the total data set (363 sentence productions) was randomly selected according to a computer-generated random number list and were reanalyzed by the original judge (intrajudge) and a second trained judge (interjudge) to obtain reliability estimates for the dependent variables of speech rate and pitch variation. Discrepancies between the reanalyzed data and the original data analysis are reported in terms of absolute mean difference and Pearson's correlation coefficients to reveal the degree of association between the data sets. Discrepancies between the reanalyzed data and the original data revealed that agreement of speech rate and pitch variation measures was high (all correlations $r > .90$), with only minor absolute differences.

Statistical Analysis

The primary goal of data analysis was to identify whether two metrics associated with speech rhythm (speech rate and pitch variation) in heard speech (exposure productions) could robustly influence the same metrics in subsequently produced speech (response production) on both an individual and group basis. Toward this end, ANOVA was used to estimate and compare group means. The normality of the residuals was determined by visual inspection of normal probability plots.

Results

Analysis of Variance

Figure 3 reflects the average speech rate for exposure and response productions, pooled across all 29 interactants, with 95% confidence intervals. The mean speech rates by speech type were habitual ($M = 4.81$), control ($M = 5.15$), hypokinetic ($M = 5.39$), and ataxic ($M = 4.46$). A repeated-measures one-way ANOVA was performed on the response productions to examine the effect of condition. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 66.69$, $p < .001$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($E = .89$). The results showed a significant effect of condition, $F(2.66, 768.87) = 40.65$, $p < .001$, $\eta^2 = .75$. Post hoc tests, using Bonferroni correction, revealed that all comparisons were significant. Thus, there was a significant difference between the habitual productions and the response productions associated with all other conditions—control speech, $t(868) = 8.51$, $p < .001$, $d = .59$, hypokinetic speech, $t(578) = 11.82$, $p < .001$, $d = .88$, and ataxic speech, $t(578) = 7.00$, $p < .001$, $d = .59$; a significant difference between control response productions and response productions associated with hypokinetic speech, $t(868) = 5.76$, $p < .001$, $d = .41$, and ataxic speech, $t(868) = 16.74$, $p < .001$, $d = 1.05$; and a significant difference between the response productions associated with hypokinetic and ataxic speech, $t(578) = 18.20$, $p < .001$, $d = 1.21$. Paired t tests were used to examine the significance of within-condition discrepancies in exposure and response productions. A significant difference was evident between the exposure and response productions associated with the degraded productions—hypokinetic speech, $t(289) = 6.93$, $p < .001$, $d = .41$, and ataxic speech, $t(289) = 40.02$, $p < .001$, $d = 2.35$. There was no significant difference between the exposure and response productions associated with the control speech.

Taken together, the statistical analyses revealed that interactants modified their speech rate to more closely align with these same properties in the exposure productions, even when speech was disordered. Interactants increased their speech rate in response to the slightly faster rates that characterized the control speech. Interactants also increased their speech rate in response to the faster rates that characterized hypokinetic speech and decreased their speech rate in response to the slower rates that characterized ataxic speech. Speech rate modifications following exposure to disordered speech did not, however, equal the abnormally fast and abnormally slow speech rates associated with hypokinetic and ataxic speech, respectively.

Figure 4 reflects the average pitch variation for exposure and response productions, pooled across all 29 interactants, with 95% confidence intervals. The mean pitch variation by speech type were habitual (M = 67.83), control (M = 65.66), hypokinetic (M = 54.19), and ataxic (M = 54.50). A repeated-measures one-way ANOVA was performed on the response productions to examine the effect of condition. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 249.16$, $p < .001$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (E = .62). The results showed a significant effect of condition, $F(1.87, 539.98) = 43.06$, $p < .001$, $\eta^2 = .69$. Post hoc tests, using Bonferroni correction, revealed no significant difference between the habitual productions and the response productions associated with control speech but found a significant difference between habitual productions and response productions associated with both hypokinetic, $t(578) = 71$, $p < .001$, $d = .62$, and ataxic speech, $t(578) = 7.67$, $p < .001$, $d = .61$. Accordingly, there was also a significant difference between the control response productions and the response productions associated with both hypokinetic, $t(868) = 8.61$, $p < .001$, $d = .59$, and ataxic speech, $t(868) = 8.41$, $p < .001$, $d = .58$. There was no significant difference between the response productions associated with hypokinetic and ataxic speech. Paired t tests were used to examine the significance of within-condition discrepancies in exposure and response productions. Significant differences were evident between the exposure and response productions associated with the degraded productions, hypokinetic speech, $t(289) = 32.58$, $p < .001$, $d = 1.91$, and ataxic speech, $t(289) = 36.61$, $p < .001$, $d = 2.15$. There was a small but significant difference between the exposure and response productions associated with the control speech, $t(579) = 4.13$, $p < .001$, $d = .46$.

Similar to the conclusions with the speech rate data, the statistical analyses revealed that interactants modified their pitch variation to be more closely align with these same properties in the disordered speech. Interactants decreased their pitch variation in response to the reduced pitch variation that characterized both hypokinetic and ataxic speech. Pitch variation modifications, however, did not match the abnormally low pitch variation associated with the disordered speech. Thus, findings from both the speech rate and pitch variation data reveal that interactants modified their speech in the direction of the incoming signal but did not completely assimilate the abnormal rhythmic production properties of the dysarthric speech.

Discussion and Conclusion

The purpose of the present study was to determine whether dysarthric speech, an acoustic signal characterized by disordered rhythmic properties, could induce production changes in the rhythmic speech patterns of healthy interactants. Such changes would support examination of rhythmic entrainment as a viable target of investigation in the domain of speech and language pathology. While the concept holds intuitive appeal as a source of communication breakdown, no previous data exist to support this. We used a quasi-conversational paradigm, similar to shadowing paradigms frequently used in the sociolinguistics literature (e.g., Babel, 2010, 2012; Babel & Bulatov, 2012; Goldinger, 1998; Namy, Nygaard, & Sauerteig, 2002; Shockley, Sabadini, & Fowler, 2004), to examine the influence that dysarthric speech exerts on the spoken productions of healthy interactants. This represents the initial step in evaluating the contribution of speech production disorders to rhythmic entrainment. Our results revealed that interacting with dysarthric speech, even within an interaction paradigm in which social interaction is not overt, invoked production shifts in the direction of the rhythmic characteristics of the exposure production. Interactants increased their rate of speech in response to productions from individuals with hypokinetic dysarthria (characterized by fast speech rate); decreased their rate of speech in response to productions from individuals with ataxic dysarthria (characterized by slow speech rate); and reduced their pitch variation in response to productions from individuals with hypokinetic and ataxic dysarthria (both of which were characterized by reduced pitch variation). Production shifts in speech rate and pitch variation, however, remained significantly different from corresponding properties in dysarthric speech. These results support the notion that entrainment is an operational interaction process—at the level of perceptual sensitivity to rhythmic features—even in the context of disordered speech.

The current findings also support a model of spoken language in which perception and production are fundamentally linked (Clark, 1997; O'Regan & Noë, 2002). An account of the bidirectional synergies of perception and production during spoken interaction in clinical populations does not exist, nor does a framework in which to illustrate how deficits in production and/or perception of rhythmic information may influence this dynamic relationship and impact human communication. We propose that the construct of entrainment offers a novel avenue to investigate a potential source of communication breakdown in interactions involving individuals with disordered speech. The present investigation gives rise to a host of research questions that offer direction for advancing this line of inquiry. We, therefore, devote the remainder of this discussion to presenting a subset of these questions in an effort to stimulate further research.

In the present study, we examined the effects of the presence of disordered speech on the spoken productions of healthy interactants. The next important question to address is how rhythmic entrainment with clinical populations is characterized in embodied face-to-face conversation. By using natural dialogue, we can explore entrainment as an entity of the conversational pair, a dynamic joint activity in which both interactants must coordinate their actions to succeed. In the context of dysarthria, we hypothesize that the associated motor impairments may interfere with entrainment in the following key areas: the ability of the person with dysarthria to modify speech relative to his or her healthy conversational partner and, to a smaller degree, the ability of the healthy partner to perceive and integrate abnormal rhythmic production patterns. We are presently addressing the nature of entrainment in a dialogue setting within a modeling paradigm. This approach, which draws on a translational methodology from the field of engineering, employs dynamic, runnable computational tools to quantify feature similarity between dialogue partners over the course of the interaction. The model output enables us to identify the presence or absence of production shifts—for both the person with dysarthria and his or her healthy conversational partner—along multiple perceptual/production dimensions and to consider behavior change as something constructed in dialogue, rather than an individual's preference or dispreference for a given rhythmic property.

The next important question concerns the mechanisms or drivers of entrainment in disordered settings. Entrainment necessitates that individual processes engage in an "exchange," allowing the properties of one to influence the other (Cummins, 2009). Modeling entrainment in dialogue, using the approach described above, will shed light on the types of rhythmic cues that drive entrainment in the context of dysarthria. Indeed, there is evidence that the answer to this question will be multifaceted. Cummins (2009) systematically degraded the availability of rhythmic information in prerecorded speech samples to examine the contribution of different cues to synchronization between speakers and recordings. Findings revealed that intelligibility alone was not sufficient to support synchronization; rather, speech synchrony was the result of several channels of information working together in an integrated fashion. For example, degrading fundamental frequency had no effect on synchrony when intelligibility of the speech samples remained intact, but when intelligibility was reduced, these pitch cues became critical to support synchronization. This laboratory finding raises the notion that the drivers of entrainment will likely change depending on the type and severity of dysarthria. This finding is further complicated by the fact that although cues like loudness, pitch variation, and speaking rate are robust carriers of speech rhythm, there is a multitude of other cooccurring communication cues that may also serve as drivers in entrainment (e.g., Gill, 2004; Gill & Borchers, 2004; Gill & Kawamori, 2002). Gill and colleagues, for example, coined the term *body moves* to refer to pragmatic acts of body movement (e.g., head nodding, body sway, shoulder movement) that convey salient rhythmic information during human interaction; others have shown that a number of conversational moves (e.g., requests for understanding) and speech backchannels (e.g., *uh huh*, *ok*) also afford rhythmic speech cues (e.g., Shimojima, Koiso, Swerts, & Katagiri, 1998). Necessary lines of research will be ones that explore and define the cues to speech rhythm, within a multilevel international framework, that drive entrainment in disordered settings.

Another potential driver is a speaker's estimation of listener need—that is, how a speaker incorporates deliberate modifications in speech production characteristics (e.g., slower speech rate, longer and more frequent pauses, exaggerated articulation) that are expected to benefit the listener (Cooke, King, Garnier, & Aubanel, 2014). These speaker-induced modifications have been observed with a number of different populations, including infants (e.g., Burnham, Kitamura, & Vollmer-Conna, 2002; Lindblom, Brownlee, Davis, & Moon, 1992), children with learning disabilities (Bradlow, Kraus, & Hayes, 2003), individuals with hearing impairment (e.g., Bradlow & Bent, 2002; Lam & Kitamura, 2012; Uchanski, Choi, Braida, Reed, & Durlach, 1996), and foreign speakers (e.g., Scarborough, Dmitrieva, Hall-Lew, Zhao, & Brenier, 2007; Skowronski & Harris, 2006; Uther, Knoll, & Burnham, 2007). Although the current findings point to a perceptual mechanism underlying the production shifts observed with exposure to dysarthric speech, this does not preclude additional influences of one's inferences about listener needs. Further, the contribution of listener needs becomes substantially more important to consider when entrainment is explored in embodied conversation. It may also be the case that speakers with dysarthria show a differential vulnerability to the drivers of entrainment across communicative exchanges, based on the composition of the dialogue pairs (i.e., familiar vs. unfamiliar communication partners).

From an intervention standpoint, a number of questions arise. First, it is critical to examine the consequences of entrainment, or lack thereof, in disordered settings—that is, does the degree of similarity/difference between dialogue partners predict how successful the interaction will be? Further, if entrainment is positively and significantly correlated with conversational success—as would be anticipated from existing literature—then can the drivers of entrainment be exploited for therapeutic gain? In other words, can the seemingly subconscious process of coordinating rhythmic behavior be consciously manipulated to enhance conversational success? Adank, Hagoort, and Bekkering (2010) show that deliberate imitation of an unfamiliar accent improves subsequent understanding of that type of speech. Can we train one or both interactants to better identify and/or integrate the rhythmic behaviors of their dialogue partner? A series of recent studies on perceptual learning of dysarthric speech demonstrate that healthy listeners can improve their ability to detect relevant acoustic cues available in the disordered speech signal (Borrie et al., 2012a, 2012b; Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2013). Does this learning translate to production changes, and further, is such learning possible in people with disordered speech? Knowledge of the drivers, and the breakdowns, of entrainment in disordered settings affords potential targets for remediation of deficits, for both the person with dysarthria and his or her primary communication partners.

Finally, the study of entrainment is not limited to dysarthria. Any impairment in the capacity to perceive, produce, and/or integrate rhythmic information has the potential to disrupt entrainment and negatively impact communication success. Accordingly, it is likely that a variety of communication disorders (e.g., apraxia of speech, stuttering, voice disorders, autism spectrum disorders, aphasia, hearing impairment, etc.) induce entrainment disruptions in conversation, albeit for different reasons. For example, individuals with a hearing impairment may have difficulty detecting the rhythmic properties of their conversational partner and therefore lack the necessary input to entrain. On the other hand, individuals with apraxia of speech may have no trouble detecting rhythm, but the associated motor impairments may prevent them from adjusting their own communicative actions accordingly. Rhythmic entrainment is a viable line of inquiry with speech and language impairment arising from a wide range of etiologies.

Indeed, the clinical import of studying rhythmic entrainment in the domain of speech-language pathology cannot be underestimated. Entrainment is critical for successful human interaction; disruptions can have deleterious consequences on developing and maintaining interpersonal relations and may exist as an integral component of the communication disorder itself. Accordingly, the study of entrainment may reveal an important source of communication difficulty and demise in quality of life in disordered settings. A model of entrainment with clinical populations would form the basis for a diagnostic tool to reveal and understand a source of breakdown in human interaction and, ultimately, identify treatment targets for improving conversational success. Entrainment, therefore, offers a new avenue for exploring speech and language impairment, addressing communication processes not currently explained by existing frameworks.

Acknowledgments

Primary support for this article was provided by the Neurological Foundation of New Zealand Grant 1119-WF, awarded to the first author. Support from the National Institute on Deafness and Other Communication Disorders Grant R01 DC 6859, awarded to the second author, is also acknowledged, as is the contribution from Hilda Torres, who assisted in collection and analysis of the manuscript data, and the Arizona State University Statistics Consulting Center. A small subset of the data in this study was part of a poster presentation at the Second International Conference on Cognitive Hearing Science for Communication, Linköping, Sweden, June 2013.

References

Adank P, Hagoort P, Bekkering H. Imitation improves language comprehension. Psychological Science. 2010; 21:1903–1909. [PubMed: 21135202]

Babel M. Dialect divergence and convergence in New Zealand English. Language in Society. 2010; 39:437–456.

Babel M. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. Journal of Phonetics. 2012; 40:177–189.

Babel M, Bulatov D. The role of fundamental frequency in phonetic accommodation. Language and Speech. 2012; 55(2):231–248. [PubMed: 22783633]

Bailenson JN, Yee N. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. Psychological Science. 2005; 16:814–819. [PubMed: 16181445]

Boersma, P.; Weenink, D. Praat: Doing phonetics by computer (Version 5.3.42) [Computer software]. 2013. Available from www.praat.org

Borrie SA, McAuliffe MJ, Liss JM, Kirk C, O'Beirne GA, Anderson TJ. Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. Language and Cognitive Processes. 2012a; 27:1039–1055. [PubMed: 24009401]

Borrie SA, McAuliffe MJ, Liss JM, O'Beirne GA, Anderson TJ. A follow-up investigation into the mechanisms that underlie improved recognition of dysarthric speech. The Journal of the Acoustical Society of America. 2012b; 132:EL102–EL108. [PubMed: 22894306]

Borrie SA, McAuliffe MJ, Liss JM, O'Beirne GA, Anderson TJ. The role of linguistic and indexical information in improved recognition of dysarthric speech. The Journal of the Acoustical Society of America. 2013; 133:474–482. [PubMed: 23297919]

Bradlow AR, Bent T. The clear speech effect for non-native listeners. The Journal of the Acoustical Society of America. 2002; 112:272–284. [PubMed: 12141353]

Bradlow AR, Kraus N, Hayes E. Speaking clearly for children with learning disabilities: Sentence perception in noise. Journal of Speech, Language, and Hearing Research. 2003; 46:80–97.

Branigan HP, Pickering MJ, Cleland AA. Syntactic co-ordination in dialogue. Cognition. 2000; 74:13–25.

Brennan SE, Clark HH. Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996; 22:1482–1493.

Brown M, Salverda AP, Dilley LC, Tanenhaus MK. Expectations from preceding prosody influence segmentation in online sentence processing. Psychonomic Bulletin and Review. 2011; 18:1189–1196. [PubMed: 21968925]

Buder, EH.; Edrington, JL. Conversational prosodic interactivity when one partner has aphasia.. In: Barbosa, PA.; Madureira, S.; Reis, C., editors. Speech Prosody 2008 Conference. Editora RG/CNPq; Campinas, Brazil: 2008. p. 501-504.

Burnham D, Kitamura C, Vollmer-Conna U. What's new, pussycat? On talking to babies and animals. Science. May 24.2002 296:1435. [PubMed: 12029126]

Chartrand TL, Bargh JA. The chameleon effect: The perception-behavior link and social interaction. Journal of Personality and Social Psychology. 1999; 76:893–910. [PubMed: 10402679]

Clark, A. Being there: Putting brain, body, and world together again. MIT Press; Cambridge, MA: 1997.

Cooke M, King S, Garnier M, Aubanel V. The listening talker: A review of human and algorithmic context-induced modifications of speech. Computer Speech & Language. 2014; 28(2):543–571.

Coulston, R.; Oviatt, SL.; Darves, C. Proceedings of the International Conference on Spoken Language Processing. Denver, CO.: Sep. 2002 Amplitude convergence in children's conversational speech with animated personas..

Cummins F. Rhythm as entrainment: The case for synchronous speech. Journal of Phonetics. 2009; 37(1):16–28.

Danescu-Niculescu-Mizil, C.; Gamon, M.; Dumais, S. Proceedings of the 20th International Conference on World Wide Web (WWW '11). ACM; New York, NY: 2011. Mark my words! Linguistic style accommodation in social media.; p. 745-754.

Furuyama, N.; Hayashi, K.; Mishima, H. Interpersonal coordination among articulations, gesticulations, and breathing movements: A case of articulation of /a/ and flexion of the wrist.. In: Heft, H.; Marsh, KL., editors. Studies in perception and action. Erlbaum; Mahwah, NJ: 2005. p. 45-48.

Garrod S, Anderson A. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. Cognition. 1987; 27:181–218. [PubMed: 3691025]

Giles H. Accent mobility: A model and some data. Anthropological Linguistics. 1973; 15(2):87–109.

Giles, H.; Coupland, J.; Coupland, N. Accommodation theory: Communication, context, and consequences.. In: Giles, H.; Coupland, J.; Coupland, N., editors. Contexts of accommodation: Development in applied sociolinguistics. Cambridge University Press; New York, NY: 1991. p. 1-68.

Gill, SP. Body moves and tacit knowing. In. In: Gorayska, B.; Mey, JL., editors. Cognition and technology: Co-existence, convergence, and co-evolution. John Benjamins; Amsterdam, the Netherlands: 2004. p. 241-266.

Gill SP. Rhythmic synchrony and mediated interaction: Toward a framework of rhythm in embodied interaction. AI & Society. 2012; 27:111–127.

Gill SP, Borchers J. Knowledge in co-action: Social intelligence in collaborative design activity. AI & Society. 2004; 17:322–333.

Gill, SP.; Kawamori, M. Coordination of gestures in a non face-to-face setting.. In: Rector, M.; Poggi, I.; Trigo, N., editors. Gestures, meaning and use. Fundacao Fernando Pessoa; Porto, Portugal: 2002.

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105(2):251–279. [PubMed: 9577239]

Gregory SW. Analysis of fundamental frequency reveals covariation in interview partners' speech. Journal of Nonverbal Behavior. 1990; 14(4):237–251.

Kohler KJ. Rhythm in speech and language. Phonetica. 2009; 66:29–45. [PubMed: 19390229]

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. Journal of Neurophysiology. 2005; 94:1904–1911. [PubMed: 15901760]

Lam C, Kitamura C. Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. International Journal of Developmental Science. 2012; 15:212–221.

Lee, C.; Black, M.; Katsamanis, A.; Lammert, A.; Baucom, B.; Christensen, A.; Narayanan, S. Proceedings of Interspeech. Makuhari, Japan: 2010. Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples.; p. 793-796.

Levitan, R.; Gravano, A.; Wilson, L.; Benus, S.; Hirschberg, J.; Nenkova, A. Acoustic-prosodic entrainment and social behavior.. Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT '12); Montreal, Canada. 2012. p. 11-19.

Levitan, R.; Hirschberg, J. Proceedings of Interspeech. Florence, Italy: 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions.; p. 3081-3084.

Lindblom B, Brownlee S, Davis B, Moon SJ. Speech transforms. Speech Communication. 1992; 11:357–368.

Liss JM, White L, Mattys SL, Lansford K, Lotto AJ, Spitzer SM, Caviness JN. Quantifying speech rhythm abnormalities in the dysarthrias. Journal of Speech, Language, and Hearing Research. 2009; 52:1334–1352.

Local, J. Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI). Saarbrucken, Germany: 2007. Phonetic detail and the organization of talk-interaction.; p. 6-10.

Louwerse MM, Dale R, Bard EG, Jeuniaux P. Behavior matching in multimodal communication is synchronized. Cognitive Science. 2012; 36(8):1404–1426. [PubMed: 22984793]

Macrae CN, Duffy OK, Miles LK, Lawrence J. A case of hand waving: Action synchrony and person perception. Cognition. 2008; 109(1):152–156. [PubMed: 18755450]

Matarazzo JD, Wiens AN. Interviewer influence on the durations of interview silence. Journal of Experimental Research in Personality. 1967; 2:56–69.

Matarazzo JD, Weitman M, Saslow G, Wiens AN. Interviewer influence on durations of interview speech. Journal of Verbal Learning and Verbal Behavior. 1963; 1:451–458.

Mattys SL, Liss JM. On building models of spoken-word recognition: When there is as much to learn from natural “oddities” as artificial normality. Perception & Psychophysics. 2008; 70:1235–1242. [PubMed: 18927006]

McFarland DH. Respiratory markers of conversational interaction. Journal of Speech, Language, and Hearing Research. 2001; 44:128–143.

Miles LK, Nind LK, Macrae CN. The rhythm of rapport: Interpersonal synchrony and social perception. Journal of Experimental Social Psychology. 2009; 45:585–589.

Nakano T, Kato N, Kitazawa S. Lack of eyeblink entrainments in autism spectrum disorders. Neuropsychologia. 2011; 49:2784–2790. [PubMed: 21699907]

Namy LL, Nygaard LC, Sauerteig D. Gender differences in vocal accommodation: The role of perception. Journal of Language and Social Psychology. 2002; 21:422–432.

Natale M. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. Journal of Personality and Social Psychology. 1975; 32:790–804.

Nazzi T, Bertoncini J, Mehler J. Language discrimination by newborns: Toward an understanding of the role of rhythm. Journal of Experimental Psychology: Human Perception and Performance. 1998; 24:756–766. [PubMed: 9627414]

Neumann R, Stack F. Mood cognition: The automatic transfer of mood between persons. Journal of Personality and Social Psychology. 2000; 79:211–223. [PubMed: 10948975]

Niederhoffer KG, Pennebaker JW. Linguistic style matching in social interaction. Journal of Language and Social Psychology. 2002; 21(4):337–360.

O'Regan J, Noë A. A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences. 2002; 24:939–1031. [PubMed: 12239892]

Pardo JS. On phonetic convergence during conversational interaction. The Journal of the Acoustical Society of America. 2006; 119:2382–2393. [PubMed: 16642851]

Peelle JE, Gross J, Davis MH. Phrase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebral Cortex. 2013; 23:1378–1387. [PubMed: 22610394]

Peterson GE, Lehiste I. Duration of syllable nuclei in English. The Journal of the Acoustical Society of America. 1960; 32:693–703.

Phillips-Silver J, Aktipis CA, Bryant G. The ecology of entrainment: Foundations of coordinated rhythmic movement. Music Perception. 2010; 28(1):3–14. [PubMed: 21776183]

Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences. 2004; 27:169–226. [PubMed: 15595235]

Putman WB, Street RL Jr. The conception and perception of noncontent speech performance: Implications for speech accommodation theory. International Journal of Sociology of Language. 1984; 46:97–114.

Reitter D, Moore JD, Keller F. Sun R, Miyake N. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. Proceedings of the 28th Annual Conference of the Cognitive Science Society. 2006:685–690.

Richardson MJ, Marsh KL, Schmit R. Effects of visual and verbal interaction on unintentional interpersonal coordination. Journal of Experimental Psychology: Human Perception and Performance. 2005; 31:62–79. [PubMed: 15709863]

Scarborough R, Dmitrieva O, Hall-Lew L, Zhao Y, Brenier J. An acoustic study of real and imagined foreigner-directed speech. The Journal of the Acoustical Society of America. 2007; 121:3044.

Shimojima A, Koiso H, Swerts M, Katagiri Y. An informational analysis of echoic responses in dialogue. Proceedings of the 20th Annual Conference of the Cognitive Science Society. 1998:951–956.

Shockley K, Sabadini L, Fowler CA. Imitation in shadowing words. Perception & Psychophysics. 2004; 66:422–429.. [PubMed: 15283067]

Shockley K, Santana MV, Fowler CA. Mutual interpersonal postural constraints are involved in cooperative conversation. Journal of Experimental Psychology: Human Perception and Performance. 2003; 29(2):326–332. [PubMed: 12760618]

Skowronski MD, Harris JG. Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments. Speech Communication. 2006; 48:549–558.

Smith, ER. An embodied account of self-other “overlap” and its effects.. In: Semin, GR.; Smith, ER., editors. Embodied grounding: Social, cognitive, affective and neuroscientific approaches (pp. Cambridge University Press; New York, NY: 2008. p. 148-159.

Street RL Jr. Speech convergence and speech evaluation in fact-finding interviews. Human Communication Research. 1984; 11(2):139–169.

Street, RL., Jr.; Giles, H. Speech accommodation theory: A social cognitive approach to language and speech behavior.. In: Roloff, ME.; Berger, CR., editors. Social cognition and communication. Sage; Beverly Hills, CA: 1982. p. 193-226.

Todd NPM, Lee CS, O'Boyle DJ. A sensorimotor theory of temporal tracking and beat induction. Psychological Research. 2002; 66(1):26–39. [PubMed: 11963275]

Truong, KP.; Trouvain, J. Proceedings Workshop on Corpora for Research on Emotion Sentiment and Social Signals. Istanbul, Turkey: 2012. Laughter annotations in conversational speech corpora: possibilities and limitations for phonetic analysis.; p. 20-24.

Uchanski RM, Choi SS, Braida LD, Reed CM, Durlach NI. Speaking clearly for the hard of hearing: IV. Further studies of the role of speaking rate. Journal of Speech and Hearing Research. 1996; 39:494–509. [PubMed: 8783129]

Uther M, Knoll MA, Burnham D. Do you speak E-NG-LI-SH? A comparison of foreigner- and infant-directed speech. Speech Communication. 2007; 49:2–7.

Webb JT. Subject speech rates as a function of interviewer behavior. Language and Speech. 1969; 12:54–67. [PubMed: 5789296]

Webb, JT. Interview synchrony: An investigation of two speech rate measures.. In: Siegman, AW.; Pope, B., editors. Studies in dyadic communication. Pergamon; New York, NY: 1972. p. 115-133.

Weismer G. Acoustic descriptions of dysarthric speech: Perceptual correlates and physiological inferences. Seminars in Speech and Language. 1984; 5:293–314.

White L, Mattys SL. Calibrating rhythm: First language and second language studies. Journal of Phonetics. 2007; 35:501–522.

Wilson M, Wilson TP. An oscillator model of the timing of turn-taking. Psychonomic Bulletin and Review. 2005; 12:957–968. [PubMed: 16615316]

Woolhouse M, Tidhar D. Demorest SM, Morrison SJ, Campbell PS. Group dancing leads to increased person-perception. In Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC 11). 2010:605–608.

Yorkston KM. Treatment efficacy: Dysarthria. Journal of Speech and Hearing Research. 1996; 39(5):S46–S57. [PubMed: 8898266]

Yorkston KM. The degenerative dysarthrias: A window into critical clinical and research issues. Folia Phoniatrica Logopaedica. 2007; 59(3):107–117.

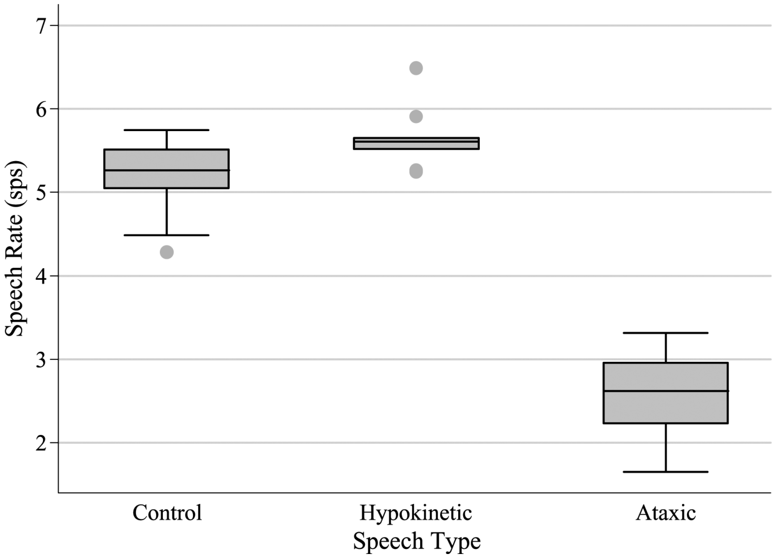

Figure 1.

Speech rate as a function of speech type for exposure productions. sps = syllables per second.

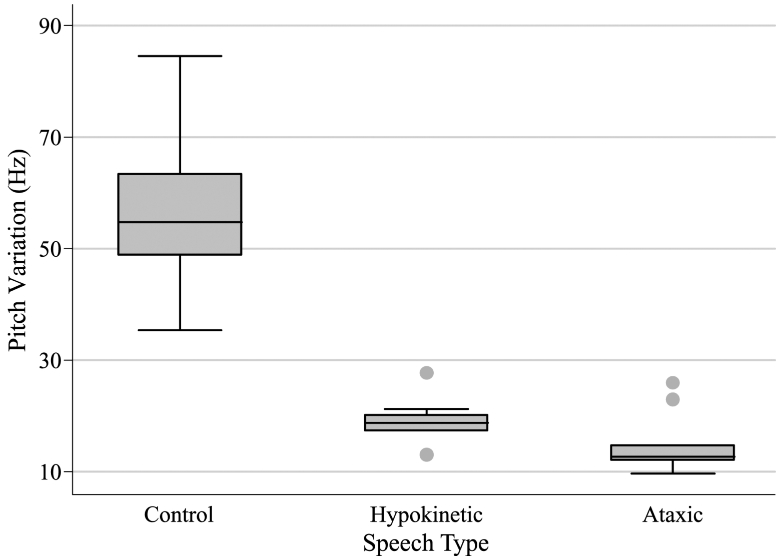

Figure 2.

Pitch variation as a function of speech type for exposure productions. Hz = Hertz.

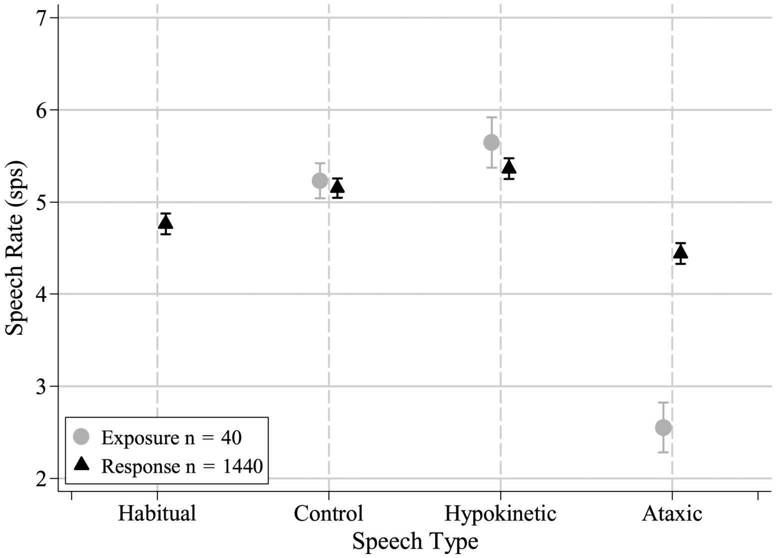

Figure 3.

Mean speech rate for exposure and response productions by experimental condition across interactants (*n* = 29).

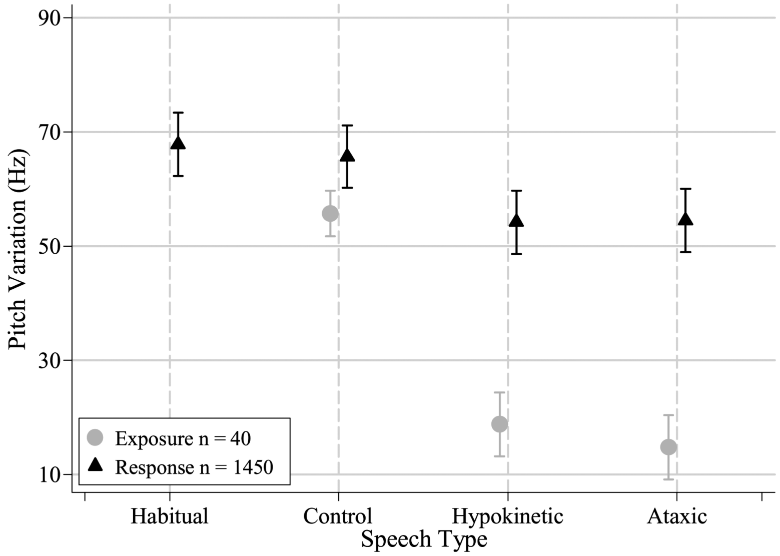

Figure 4.
Mean pitch variation for exposure and response productions by experimental condition across interactants (*n* = 29).

Appendix A

Speech Stimuli

1. The supermarket chain shut down because of poor management.
2. Much more money must be donated to make this department succeed.
3. In this famous coffee shop they serve the best doughnuts in town.
4. The chairman decided to pave over the shopping center garden.
5. The standards committee met this afternoon in an open meeting.