Applied Psycholinguistics **37** (2016), 1201–1220
doi:10.1017/S0142716415000545

Speech rates converge in scripted turn-taking conversations

BENJAMIN G. SCHULTZ and IRENA O'BRIEN
McGill University

NATALIE PHILLIPS
Concordia University

DAVID H. MCFARLAND
Université de Montréal

DEBRA TITONE and CAROLINE PALMER
McGill University

Received: November 3, 2014   Accepted for publication: May 13, 2015

ADDRESS FOR CORRESPONDENCE
Caroline Palmer, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue,
Montreal, QC H3A 1B1, Canada. E-mail: mailto:caroline.palmer@mcgill.ca

ABSTRACT
When speakers engage in conversation, acoustic features of their utterances sometimes converge. We examined how the speech rate of participants changed when a confederate spoke at fast or slow rates during readings of scripted dialogues. A beat-tracking algorithm extracted the periodic relations between stressed syllables (beats) from acoustic recordings. The mean interbeat interval (IBI) between successive stressed syllables was compared across speech rates. Participants' IBIs were smaller in the fast condition than in the slow condition; the difference between participants' and the confederate's IBIs decreased across utterances. Cross-correlational analyses demonstrated mutual influences between speakers, with greater impact of the confederate on participants' beat rates than vice versa. Beat rates converged in scripted conversations, suggesting speakers mutually entrain to one another's beat.

Humans are generally social creatures that crave acceptance. Group affiliation can be facilitated when an individual changes his or her behavior in such a way that it conforms to the norms of the group (Giles, 1977). According to the *communication accommodation theory* (Giles, 1977), conversational partners change their behaviors, such as speech or gesture, during an interaction to communicate effectively and show affiliation with a group. Specifically, conversational partners converge upon a common speech style. Previous research has shown that interlocutors within a conversation converge in terms of respiration (McFarland, 2001), pronunciation (Pardo, 2006), pitch (i.e., fundamental frequency) and syllable duration (Bosshardt, Sappok, Knipschild, & Holscher, 1997), and intensity (Huber, 2008; Natale, 1975). The convergence of speech rate within a conversation, however, has received little attention (but see Manson, Bryant, Gervais, & Kline, 2013; Street, 1984; Webb, 1969, 1972). Speech rate is of significant interest because it varies considerably during normal conversation (Miller, Grosjean, & Lomanto, 1984) and can substantially alter the acoustic information that permits listeners to discriminate one speech sound from another (Volaitis & Miller, 1992). The present study investigates the convergence of speech rate in scripted, turn-taking conversations between a confederate and participants.

SPEECH RATE CONVERGENCE AND INTERPERSONAL SYNCHRONY

Only a few papers have examined speech rate convergence, and most of these have measured speech rate as the number of syllables per second (or syllables per minute). Jungers, Palmer, and Speer (2002) showed that listeners changed their speech rate to match those of auditory recordings of a speaker who spoke at fast or slow rates. Similarly, Cummins (2009) established that speakers can adjust their speech timing to occur simultaneously with audio recordings. In natural conversation, Street, Street, and Van Kleek (1983) demonstrated that the speech rates of female children and female adults converged for three out of four children. However, the convergence of speech rate in natural conversation has remained elusive except under certain circumstances: same-sex male but not same-sex female dyads converged when instructed to imitate (Pardo, Jay, & Krauss, 2010), and those initially assigned an instructor role in a conversation converged more when they assumed a receiver role compared with those who were initially assigned the receiver role (Pardo et al., 2013). In the context of interviews, positive correlations have been demonstrated between the speech rates of interviewers and interviewees (Street, 1984; Webb, 1969, 1972). Taken together, these results suggest that speech rate can converge between two speakers, but what is the cognitive mechanism that underlies speech rate convergence?

Wilson and Wilson (2005) suggested that a similar mechanism underlies speech rate convergence in turn-taking conversation and interpersonal coordination when moving to a common beat. The beat, a percept, is a regular temporal pulse that can be abstracted from a signal based on salient features, such as changes in pitch and intensity (London, 2012; Povel & Essens, 1985). It is arguable that the most pervasive situation in which the beat is perceived, and subsequently acted upon, is spontaneous synchronization to music and dance. For example, Janata, Tomic, and Haberman (2012) found that people spontaneously tap their foot or bob their head while listening to music and that the prevalence of these movements relates to how strongly the music implies a beat. People also spontaneously coordinate movements with one another when performing periodic actions, such as pendulum swinging (Schmidt, Richardson, Arsenault, & Galantucci, 2007) and rocking in rocking chairs (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007). In speech, listeners perceived regular beats in recordings of English and French utterances, and both English and French speakers perceived greater regularity for English compared to French utterances (Lidji, Palmer, Peretz, & Morningstar, 2011). Therefore, listeners in a conversational dynamic may extract beats from a speaker's vocal productions in a similar way to that shown in music and, consequently, the perceived beats may affect their vocal productions when it is their turn to speak. It follows that the same mechanisms could be involved in both speech rate convergence and beat synchronization.

A mechanism for how people synchronize to the beat is offered by the *dynamic attending theory* (Jones, 2009; Large & Jones, 1999), which describes expectancy as sinusoidal neural oscillations that synchronize with, and adapt to, temporal regularities in an external signal. The process by which neural oscillations synchronize with an external beat is called *entrainment*, and there is some evidence that this occurs during the perception of nonlinguistic auditory rhythmic sequences (Nozaradan, Peretz, & Mouraux, 2011, 2012). Wilson and Wilson (2005) propose that entrainment is also responsible for speech rate convergence within a conversation. Coordination of the EEG signals of speakers and listeners has been demonstrated when listeners are asked to focus on one of two superimposed audiovisual human speakers (Kuhlen, Allefeld, & Haynes, 2012). While entrainment of neural oscillations was not examined per se, there was evidence that listeners' EEG signals were correlated with the speaker's EEG signal at multiple lags hypothesized to correspond to multidimensional information conveyed by the speaker; regularities in speech timing might be one such dimension in which speakers and listeners coordinate.

For turn-taking conversations, the dynamic attending theory would posit that the neural oscillations of the interlocutors would become coupled: the listener would entrain to the speech rate of the speaker's vocal productions. Consequently, when it is the listener's turn to speak, the rate of the listener's vocal productions would be influenced by the speech rate of the previous speaker. Through turn taking, interlocutors would be given the opportunity to entrain to each other's speech rate and should demonstrate convergence. Thus, there would be a bidirectional relationship where each of the interlocutors is affecting the other's speech rate. Based on the dynamic attending theory, it was expected that a confederate's speech rate would influence that of participants and vice versa.

Wilson and Wilson (2005) claim that the syllable is the most likely basis for entrainment in speech across languages due to differences in stress patterns between stress-timed, syllable-timed, and Mora-timed speech. Although there is evidence to support differences between stress-timed and syllable-timed languages (e.g., Bertinetto & Fowler, 1989; Ramus & Mehler, 1999), several authors have argued that these classifications are not as distinct as once thought (e.g., Arvaniti, 2012; Dauer, 1983; for a review, see Cummins, 2012). Regardless of how distinct these categories are, all languages contain patterns of acoustic stress arising from changes in duration, intensity, and frequency in the speech signal. Here, we propose that people entrain to speech in the same way as to music, that is, not to each syllable or note onset but instead to regular patterns of intensity and frequency that are perceived as speech stress. To measure the beat of speech, one must consider both the patterns of syllable onsets and the patterns of acoustic stress. We suggest a measure of speech rate that is informed by, but not identical to, syllable onset times. Specifically, we propose that salient syllables, perceived as beats, are a unit of entrainment for speech rate (see Classé, 1939; Cummins & Port, 1998).

A broad definition of beat that applies to music and speech is a regular, but not necessarily isochronous, pulse that an individual perceives as aligning with accented or stressed points in time. The beat is related to, but different from, rhythm and prosody. Rhythm is defined as the "systematic patterning of sound in terms of timing, accent, and grouping" (Patel, 2008, p. 96). Similarly, prosody has been defined as perceived "stress, rhythm, and intonation in spoken sentences" (Kjelgaard & Speer, 1999, p. 153). The beat is a regular pulse that is induced by rhythm and prosody and the acoustic features that are perceived as stress. Speech rate has been measured in several ways. Syllables per second (and words per minute) measures the ratio of syllables (or words) spoken relative to the duration of the utterance. This calculation can be affected by other factors, such as midutterance pauses and speech errors, and does not account for acoustic stress patterns in speech that might induce a sense of beat. Articulation rate, similar to syllables per second in that it measures the rate of syllable onsets within an utterance while eliminating pauses and silences (Miller et al., 1984), is another common measure of speech rate. However, eliminating pauses might remove meaningful temporal information that contributes to the speech rate and prosody of speakers. Articulation rate, pause duration, and the frequency of pauses have all been found to contribute to the perception of speech rate (Grosjean & Lane, 1976; Miller & Grosjean, 1981), although articulation rate appears to play a larger role. Like syllables per second, however, articulation rate does not consider acoustic stress patterns. The perceptual center (Morton, Marcus, & Frankish, 1976), defined as the perceptual onset of syllables in relation to acoustic features, has also been used to examine the speech rate of individuals in speech cycling tasks (Cummins & Port, 1998). Perceptual centers have been hypothesized to indicate the beat or regularity in speech sequences that arise as a result of a combination of acoustic properties. However, perceptual centers often do not pertain to a consistent set of acoustic features or stress patterns (Marcus, 1981), are difficult to measure reliably (see Villing, Ward, & Timoney, 2003), and have not been applied to measure speech rate in conversational speech. To account for the contributions of syllable timing, pauses, and acoustic stress patterns to speech rate, we employed another method to unify these factors that are present in acoustic speech signals. Here, we apply an automatic computational algorithm for beat tracking that was originally developed for music (Ellis, 2007a) to acoustic speech recordings to measure the beat rate of speech.

BEAT-TRACKING ALGORITHMS

Beat-tracking algorithms aim to find beat onsets and the beat rate of an acoustic signal. Such algorithms are more often applied to music than to speech recordings (see McKinney, Moelants, Davies, & Klapuri, 2007), and this is likely because music tends to conform to a regular, perceptually isochronous beat and lends itself to beat extraction more easily than speech. Speech is temporally dynamic and contains acoustic cues (e.g., stress) that form temporal intervals between stressed syllables, which we term interbeat intervals (IBIs). Thus, successful beat tracking within a speech signal must be sensitive to acoustic indicators of stress and allow for high temporal variability. The present study employs a beat-tracking algorithm (Ellis, 2007a) that permits nonisochronous beats and detects beats based on changes in pitch and intensity that are analogous to speech stress (Sluijter, Van Heuven, & Pacilly, 1997).

The beat-tracking algorithm was selected for three reasons. First, the algorithm has been shown to correspond to the perceived beat for musical stimuli (McKinney et al., 2007) and can be applied to acoustic data. Second, the algorithm identifies possible beat locations by weighting perceptual intensity at different frequency bands (i.e., Mel frequencies; see Stevens, Volkmann, & Newman, 1937) and assigning beat strength based on perceptual loudness, a correlate of speech stress. Third, the algorithm allows for nonisochronous beats. There is evidence that temporal intervals between stressed syllables are rarely isochronous (Classé, 1939; Crystal & House, 1990; Dauer, 1983; Jassem, Hill, & Witten, 1984; Shen & Peterson, 1962; for reviews, see Cutler, 1991; and Lehiste, 1977). Several of these studies have used foot duration or syllable duration to indicate speech stress. However, there is evidence that a range of acoustic factors, including intensity and frequency, also influence the perception of stress (Sluijter et al., 1997). The current study does not make any claims as to whether stress patterns are isochronous; as demonstrated by Lidji et al. (2011), listeners are able to tap regularly to a perceived beat for both English and French speech, indicating that strict isochrony is not necessary to perceive a regular beat. Moreover, Large, Fink, and Kelso (2002), have shown that people quickly adapt to temporal perturbations in isochronous and nonisochronous rhythms. The present study regards the intervals between beats in speech as dynamic, nonisochronous, temporal intervals between successive stressed syllables that may imply a global beat rate while tolerating local tempo changes or temporal perturbations.

The algorithmic beat tracker offers several advantages over traditional speech rate measures. First, the algorithm reveals temporal regularities between stressed syllable onsets at local levels and does not consider every syllable to contribute to the beat. Second, the beat continues to cycle through periods of silence (e.g., midutterance pauses) if there is enough evidence for the beat's continuation from surrounding stressed syllable onsets, consistent with evidence that suggests the perception of a beat continues briefly when a speech stream stops (Wilson & Wilson, 2005) or when a musical stream contains silent intervals (Nozaradan et al., 2012; Povel & Essens, 1985). We refer to *speech rate* as speech timing at the syllable level and *beat rate* as speech timing at the stress level (as identified by the beat-tracking algorithm).

HYPOTHESES

We examined how participants' beat rate was influenced by the beat rate of a confederate when the confederate spoke at fast or slow speech rates (as cued by syllable rate) during readings of scripted dialogues from two plays. The mean IBI between successive stressed syllables, based on the beat-extraction algorithm, was compared between the fast and slow confederate speech rate conditions. Three hypotheses were tested: (a) participants' beat rates are faster when the confederate speaks at a fast rate compared to a slow rate, (b) the beat rates of the participant and the confederate converge over the duration of the dialogue, and (c) the beat rate of the confederate has a greater influence on the beat rate of the participant than vice versa. The rationale behind the third hypothesis is that the confederate is purposefully driving a speech rate and taking a leader role in the experimental manipulation. This does not preclude that the beat rate of the participant will also influence the beat rate of the confederate; based on mutual entrainment (Wilson & Wilson, 2005), it was also hypothesized that the participants' beat rates influence those of the confederate.

METHOD

Participants

The participants (N = 38) were students recruited from McGill University, and they were compensated with course credit or a small payment. They met the criteria of speaking North American English as their first language, having no self-reported speech impairments, and having normal hearing as verified by pure-tone audiological screening. These participants had a mean age of 20.7 years (SD = 2.48, range = 18–30 years), and 35 participants were female (3 male). The confederate was a 21-year-old female undergraduate student of the University of McGill who spoke North American English as her first language.

Materials and equipment

Three different scripts were read by each participant. The first script was a portion of the consent form containing 18 utterances (454 syllables total) that was read aloud to measure each individual's natural (i.e., uncued) beat rate. Participants read the consent form aloud once prior to the experimental trials; the confederate performed this task once before testing commenced and prior to being exposed to the prompts for speech rate conditions. The dialogue scripts were taken from *Death of a Salesman*, written by Arthur Miller (1949), which contained 33 utterances for the confederate and 32 utterances for the participant, and *The Importance of Being Earnest*, written by Oscar Wilde (1908; originally performed in 1895), which contained 27 utterances for both the confederate and the participant. Scripted dialogues were used for two reasons: first, they provide a controlled situation for turn taking in which the participant and the confederate produce a known number of syllables per utterance; and second, the use of scripts reduces the effect of potential confounds such as amount of time required for planning speech or comprehending what was said by the partner. It has been shown that increased speech rate can result in decreased comprehension (Krause & Braida, 2002). Thus, the ability to read the scripted dialogue reduces the likelihood that differences between confederate speech rate conditions (fast or slow) are due to differences in the level of comprehension. All vocalizations were recorded using two AKG C420 head-mounted microphones to two separate channels (one for the confederate's speech, and the other for the participants' speech), and saved onto a Marantz portable solid-state disk (PMD670) at a sample rate of 44.1 kHz.

Design and procedure

Informed consent was obtained. Participants then completed the following tasks. Participants read a portion of the consent form aloud and their speech was recorded. This task lasted about 1.5 min and allowed participants to become comfortable with the microphones. Then, the participant was introduced to the confederate under the guise that the confederate was another participant. The participant then read the two scripted dialogues with the confederate under conditions of either fast confederate speech rate (prompted with a metronome IBI of 0.17) or slow confederate speech rate (prompted with a metronome IBI of 0.33). Participants read each script once and the order of the dialogue scripts (*Death of a Salesman* and *The Importance of Being Earnest*) was counterbalanced across participants; one half received *Death of a Salesman* in the fast condition and *The Importance of Being Earnest* in the slow condition, and the other half received the opposite combination. In a separate room, the confederate was prompted (via headphones) by a metronome at each of the speech rates prior to speaking with the participant; the metronome did not sound during any trial. The order of speech rate conditions was counterbalanced, and to provide a cover story as to why the confederate was speaking at different rates in the two dialogues, the confederate entered (or reentered) the room with a caffeinated energy drink for the fast condition. None of the participants were aware of the confederate's status as indicated by verbal questioning of the participant at the end of the experiment. The confederate was not informed of the experiment hypotheses, and she reported that she maintained the same speech rate throughout each trial. Moreover, the confederate was not informed that acoustic speech recordings would be subjected to beat-tracking analyses.

RESULTS

Analysis

Each utterance was measured for speech rate using both traditional (syllables per second) and beat rate using the beat-tracking (algorithmic) analyses; we describe the algorithm measurements first. The beat-tracking algorithm, written for MatLab (Mathworks, Natick, MA), is available online (Ellis, 2007a, 2007b). The version of the algorithm used in the present experiment contains additional parameters, described here, than those in the most recent publication (Ellis, 2007a).[1] The primary input for the beat-tracking algorithm is an acoustic signal (see Figure 1a for an example) in the form of a waveform. There are also two optional parameters of initial beats per minute and tightness. Initial beats per minute refers to the starting or expected tempo of the acoustic excerpt, and reflects a prescribed tempo that the algorithm may or may not maintain depending on whether the beat tracker finds any evidence for the suggested rate. The tightness parameter reflects the degree to which the global tempo is allowed to change throughout the utterance. Lower tightness values (e.g., 1) allow more dynamic IBIs and more change in the algorithm's beat estimation within an utterance, and higher tightness values (e.g., 600) restrict the IBIs and allow less change in the algorithm's beat estimation. A low tightness value of 10 (highly dynamic) was implemented here. The initial

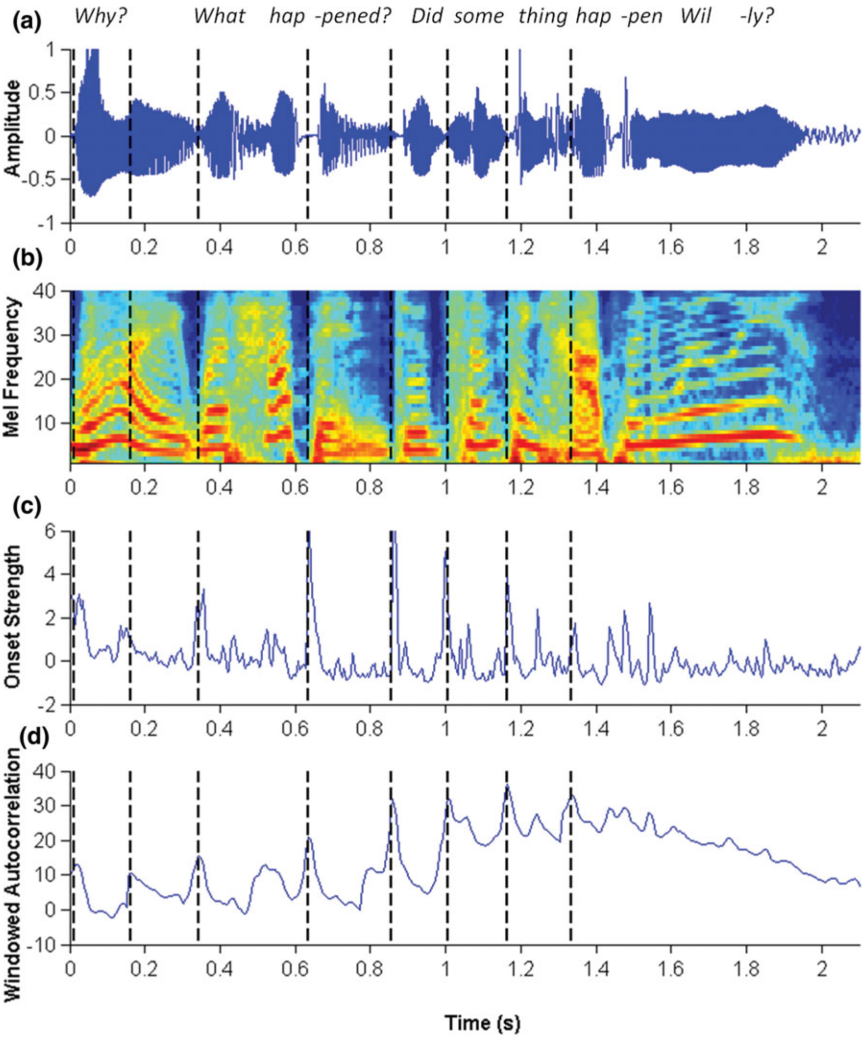

Figure 1. Example of tempo calculation in Ellis' (2007a) beat-tracking algorithm. Beats are indicated by dashed vertical lines. Interbeat intervals correspond to the temporal intervals between successive beats based on the periodic moments of high onset strength (see Onset Strength panel). In order, these interberat intervals are 0.15, 0.18, 0.29, 0.22, 0.15, 0.16, and 0.17 s. The example is from *Death of a Salesman*, first confederate utterance, "Why? What happened? Did something happen, Willy?"

beats per minute was set so that the prescribed tempo of each utterance was the mean beat rate of the previous utterance by the same speaker. However, the dynamic tightness level allowed the prescribed tempo to be ignored if there was any deviation from the prescribed beats per minute. The prescribed beats per

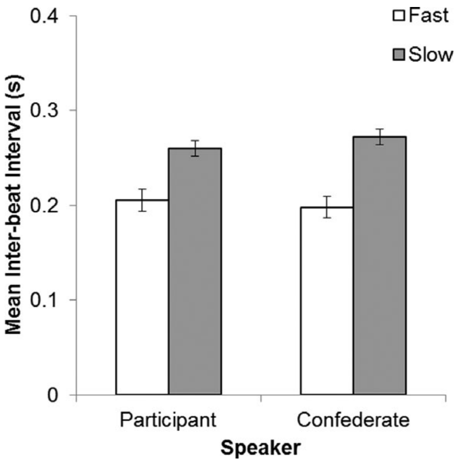

Figure 2. Mean interbeat interval for participants and confederate in the (white bars) fast and (gray bars) slow speech rate conditions. Whiskers represent the standard error of the mean.

minute was used to prevent the beat tracker from providing multiples of the IBI, which could occur when using low tightness values on short utterances.

The beat-tracking algorithm calculates beat onset times through several stages. First, the acoustic signal is transformed into Mel frequencies (Stevens et al., 1937; see Figure 1b) that correspond to the subjective experience of loudness at perceptually salient frequencies for adult humans (~20–16,000 Hz) using a crude perceptual model (for full details, see Ellis, 2007a). Second, onsets strengths are calculated by summing intensity across all Mel frequencies for each point of time (see Figure 1c). Third, windowed (Figure 1d) autocorrelations (32-ms window, 4-ms time step) are conducted at multiple time points to find periodic recurrence of moments of high onset strength. If multiple autocorrelations are significant, then the algorithm will select the autocorrelations based on the initial beats per minute and tightness. The output is a series of beat onsets for the acoustic signal (see dashed lines in Figure 2) from which the IBIs are calculated by subtracting the time of the current beat onset (t^n) from the time of the next beat onset (t^{n+1}) as shown in Equation 1. The present study applied the beat-tracking algorithm to each utterance spoken by the participant and confederate individually. The mean IBI for each utterance spoken by the participant and the confederate was used as the primary measure of beat rate within each utterance.

$$\mathrm{IBI}^n = t^{n+1} - t^n. \tag{1}$$

Mean IBI by speaker and speech rate

The baseline measures of beat rate from the consent form readings showed that the confederate had a mean IBI of 0.22 s (*SEM* = 0.07), and participants had a mean IBI of 0.23 s (*SEM* = 0.06, range = 0.1–0.37), as determined by the beat-tracking algorithm. The range of the participants' baseline beat rates encompassed the mean IBI of the confederate and the prescribed metronome beat in both the fast (0.17 s) and slow (0.33 s) speech rate conditions. Therefore, if convergence is observed in both the fast and the slow conditions, then participants would have changed their beat rate (from baseline) in at least one of those conditions.

To test the hypothesis that participants' dialogue beat rates are faster when the confederate spoke at a fast rate compared to a slow rate in scripted conversations, mean IBI values (collapsed across utterances) were compared between fast and slow confederate speech rate conditions. Paired-samples t tests revealed significant differences between fast and slow rates for the participants' beat rate, t (37) = –4.09, $p < .001$, $d = 0.24$, and the confederate's beat rate, t (37) = –5.25, $p < .001$, $d = 0.30$. As shown in Figure 2, IBIs were smaller (i.e., the rate was faster) for the fast condition compared to the slow condition for both the participant and the confederate. These results support the hypothesis that participants' beat rate was affected by the rate at which the confederate was speaking (fast or slow), and confirm that the confederate spoke at a faster rate in the fast condition than in the slow condition.

To examine the degree to which participants deviated from their baseline beat rates, the absolute difference between baseline IBIs and IBIs in the fast and slow speech rate conditions were both analyzed. The absolute difference was used because, due to the range of baseline IBIs, some participants increased their rate in both conditions in response to the confederate, some participants decreased their rate in both conditions, and others increased their rate in the fast condition and decreased their rate in the slow condition. One-sample t tests were conducted to assess whether absolute differences were significantly greater than zero. Participants demonstrated significant deviations from their baseline speech rates in the fast, $M = 0.10$, $SEM = 0.01$, t (37) = 7.62, $p < .001$, $d = 0.35$, and slow, $M = 0.08$, $SEM = 0.01$, t (37) = 8.26, $p < .001$, $d = 0.33$, conditions. These results indicate that participants deviated from their baseline beat rate in the fast and slow speech rate conditions, further supporting the hypothesis that participants' beat rates are affected by the confederate speech rate conditions.

Beat rate convergence across utterances

To test the hypothesis that the beat rates of participants and the confederate converged during the dialogues, we calculated the absolute difference between the IBIs of the participant and the confederate for each pair of utterances (i.e., the difference between a confederate's utterance and the participant's immediately following utterance). Because the two scripts contained a different number of utterances, the mean absolute IBI difference was calculated for the first (start), middle, and final (end) eight utterances (see Figure 3). Absolute IBI differences were analyzed using a 2 (rate; fast, slow) × 3 (section; start, middle, end) repeated measures analysis of variance. There were significant main effects of rate, F (1, 37) = 12.55, $p = 0.001$, $\eta_p^2 = 0.25$, and section, F (2, 74) = 5.35, $p = .007$, $\eta_p^2 = 0.13$, but no significant interaction between rate and section, F (2, 74) = 1.15, $p = .32$, $\eta_p^2 = 0.03$. As shown in Figure 3, the significant main effect of rate indicated that the difference between the rates of participants and the confederate

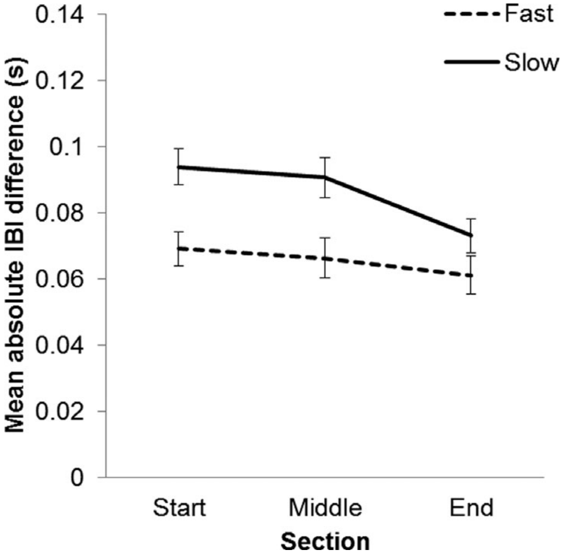

Figure 3. Mean absolute (unsigned) interbeat interval difference between the confederate and participant for the start, middle, and end sections by (dashed line) fast and (solid line) slow speech rate conditions. Whiskers represent standard error of the mean.

was smaller in the fast condition than in the slow condition. Pairwise comparisons revealed that absolute IBI differences at the start were not significantly different from those in the middle ($p = .53$) but were significantly greater than those at the end ($p = .008$). Absolute IBI differences in the middle were also significantly greater than those at the end ($p = .006$). A linear trend analysis confirmed that absolute IBI differences decreased over sections, F (2, 74) = 7.89, $p = .008$, $\eta_p^2 = 0.18$. These results indicate that the beat rates of participants and the confederate converged over the course of each trial. Thus, the hypothesis that the beat rate of participants and the confederate converges during the dialogue was supported.

Cross-correlations between confederate's and participants' IBIs

To further examine the joint influence of the participants' and confederate's beat rates across the dialogues, cross-correlations were conducted on participant and confederate IBIs at lags 1, 0, and –1. The confederate had the first utterance in each dialogue. Lag 0 was defined as the cross-correlation of the confederate's *n*th utterance with the participant's *n*th utterance (which immediately followed). A positive lag 0 value indicates the participant adjusted his or her beat rate to the beat rate of the previous utterance of the confederate. Lag +1 was defined as the correlation of the confederate's *n*th utterance with the participant's *n*th +1 utterance. A positive lag +1 value indicates that the beat rate of the participant's

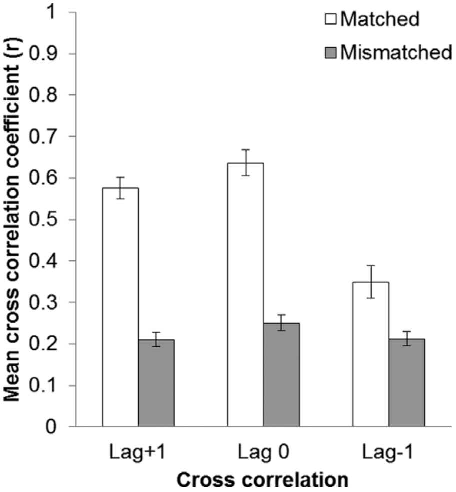

Figure 4. The mean cross-correlations coefficients (*r*) at lag +1, lag 0, and lag –1 between (white bars) the interbeat intervals of the participant and (gray bars) the within-trial confederate and the mean upper confidence interval of the cross-correlation coefficients between the interbeat intervals of each participant paired with all other-trial confederates. Mismatched bars represent chance performance estimated by Jackknifing techniques. Whiskers represent the standard error of the mean.

utterances was influenced by the beat rate of the confederate two utterances previous. Lag –1 was defined as the correlation of the confederate's *n*th utterance with the participant's *n*th – 1 utterance. A positive lag –1 value indicates that the confederate's beat rate is being influenced by the beat rate of the previous utterance spoken by the participant.

A repeated measures analysis of variance on the cross-correlations by rate (fast or slow) and lag (lag +1, lag 0, or lag –1) revealed a significant main effect of lag, $F\,(2, 74) = 29.10, p < .001, \eta_p^2 = 0.44$, but no significant main effect of rate, $F\,(1, 37) = 0.66, p = .42, \eta_p^2 = 0.02$, or interaction between rate and lag, $F\,(2, 74) = 0.67, p = .51, \eta_p^2 = 0.02$. Pairwise comparisons (using the Bonferroni adjustment) between lags revealed significant differences between lag +1 and lag –1 ($p < .001$), lag 0 and lag –1 ($p < .001$), and lag +1 and lag 0 ($p < .001$). Figure 4 indicates that lag 0 demonstrated the largest positive cross-correlations, followed by lag +1, and that lag –1 demonstrated significantly smaller cross-correlation coefficients than lag +1 and lag 0. These results indicate that the confederate's beat rates influenced those of participants more than the participants' beat rates influenced those of the confederate. Thus, the hypothesis that the beat rate of the confederate has a greater influence on the beat rate of the participant than vice versa was supported.

To determine whether the cross-correlation coefficients were significantly different from those expected by chance, Jackknife resampling simulations (from the methods of Quenouille, 1949; Tukey, 1958) were conducted on cross-correlation coefficients. The time series for each participant (P_x) was cross-correlated with the time series of the mismatched confederate utterances ($C_{1,\ldots,N}$, excluding C_x) in all other trials within the same speech rate and script condition ($N – 1 = 18$ mismatched comparisons). The means and confidence intervals (CI) of the *r* values were extracted from the 18 mismatched pairings per condition for lag +1, lag 0, and lag –1. Due to the expectation that cross-correlation coefficients between the participant and the confederate in the matched pairings are positive, we compared the cross-correlation coefficients of the matched pairs with the upper CI of the mismatched pairs. This is a more conservative test because the upper CIs are naturally larger than the mean values.

For each lag (lag +1, lag 0, or lag –1), paired-samples *t* tests were conducted between the matched pairs and the upper CI of the cross-correlation coefficients for the mismatched pairs. As shown in Figure 4, cross-correlation coefficients for matched pairs were larger than the upper CI of cross-correlation coefficients for mismatched pairs for lag +1, $t\ (75) = 30.45$, $p < .001$, $d = 3.39$; lag 0, $t\ (75) = 54.85$, $p < .001$, $d = 6.86$; and lag –1, $t\ (75) = 21.97$, $p < .001$, $d = 2.63$. The Jackknife simulations were repeated by holding constant each confederate time series and pairing it with that from every other mismatched participant, and yielded an identical pattern of results. Taken together, these results showed that cross-correlation coefficients for lags +1, 0, and –1 were all significantly greater than levels expected by chance. This indicated that the beat rate of the confederate significantly influenced those of the participants (lags 0 and +1) and that the beat rates of the participants also significantly influenced that of the confederate (lag –1). Thus, there was a bidirectional influence of beat rate between the confederate and participants.

Syllables per second analyses

The same analyses were conducted with syllables per second as the dependent variable. The number of syllables per utterance and the utterance durations were assessed manually. Table 1 shows the speech rate values for each condition. For the consent form speech rate, the confederate spoke at 5.14 syllables per second, and participants had a mean of 5.65 syllables per second ($SEM = 0.07$, range = 4.6–96.62). For overall speech rate in dialogues (see Table 1), both participants, $t\ (37) = 6.10$, $p < .001$, $d = 0.58$, and the confederate, $t\ (37) = 39.99$, $p < .001$, $d = 4.33$, demonstrated significantly more syllables per second in the fast condition than in the slow condition. One-sample *t* tests were conducted on the absolute difference between baseline speech rates and condition speech rates to assess whether speech rates in the fast and slow conditions differed from baseline speech rates in the consent form reading. Participants demonstrated significant deviations from their baseline speech rates in both the fast, $M = 0.57$, $SEM = 0.12$, $t\ (37) = 4.67$, $p < .001$, $d = 0.66$, and slow, $M = 1.00$, $SEM = 0.17$, $t\ (37) = 5.94$, $p < .001$, $d = 0.98$, conditions.

For the rate convergence analysis, the difference between the participants and the confederate significantly decreased over sections for the fast condition, linear $F\ (2, 74) = 39.56$, $p < .001$, $\eta_p^2 = 0.52$, but not for the slow condition, linear $F\ (2, 74) = 0.01$, $p = .94$, $\eta_p^2 = 0.00$. As shown in Table 1, this was due to

Table 1. *Means and standard errors of the mean for syllables per second measures in dialogues*

		Fast		Slow	
Dependent Variable	Independent Variable	Mean	*SEM*	Mean	*SEM*
Mean syllables per second	Participant	5.26	0.07	4.89	0.06
	Confederate	5.58	0.05	3.42	0.03
Absolute syllables per second difference	Start	1.22	0.05	1.67	0.08
	Middle	0.96	0.04	1.44	0.05
	End	0.82	0.04	1.66	0.06
Cross-correlation coefficients (*r*)	Lag 1	0.64	0.01	0.64	0.01
	Lag 0	0.78	0.01	0.79	0.01
	Lag −1	0.61	0.01	0.61	0.01

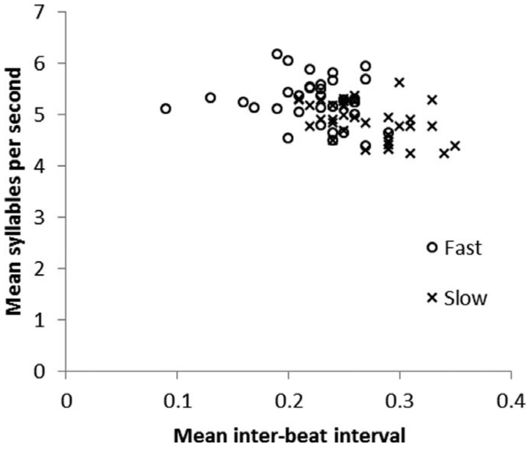

Figure 5. Scatterplots demonstrating the relationships between mean interbeat interval and mean syllables per second for participants in the (○) fast and (×) slow speech rate conditions.

an increase in the absolute difference in the end section for the slow condition. For cross-correlational analyses (see Table 1), pairwise comparisons showed that cross-correlations at lag 0 were greater than at lag 1 and lag –1 (*p*s $< .001$), and cross-correlations at lag 1 were greater than at lag –1 ($p < .001$). With the exception of the rate convergence analysis for the slow condition, analyses on syllables per second yielded identical trends to those demonstrated by the beat tracker analyses. Overall, significant negative correlations between IBIs and syllables per second measures were demonstrated for the participants, r (74) = –.35, p = .002, and confederate, r (74) = –0.49, $p < .001$, suggesting that the two measures are related but not identical (see Figure 5).[2] These relatively small correlation values indicate that the beat tracker is measuring a slightly different aspect of speech rate, that is, the rate of quasiregular positions of speech stress.

DISCUSSION

This study examined the influence of a confederate's beat rate on the beat rate of naïve participants during the reading of dialogues. Three major findings addressed predictions based on convergence between speakers. First, participants' beat rates were faster when the confederate spoke at a faster rate than at a slower rate. Second, beat rate convergence increased over the course of each dialogue. Third, beat rate influences were bidirectional, but the confederate had a greater influence on the beat rate of the participant than vice versa. Overall, findings confirmed that the beat rate of speech, as measured by the beat-tracking algorithm (Ellis, 2007a), converged when participants read a conversational dialogue with a confederate. Analyses using the traditional speech rate measure, syllables per second, generally corresponded with results elicited by the beat tracker. These results complement previous research that demonstrates that interlocutors adjust acoustic parameters of speech productions (e.g., pitch, duration, and intensity) when conversing with a partner (Bosshardt et al., 1997; Huber, 2008; Natale, 1975; Pardo, 2006). Findings are also in line with research that has shown convergence of respiration (McFarland, 2001), gesture (Giles, 1977), as well as vocabulary, posture, linguistic style, facial expressions, head nods, and yawns (see Branigan, Pickering, Pearson, & McLean, 2010; Burgoon, Stern, & Dillman, 1995; Dale, Fusaroli, Duran, & Richardson, 2013; Gill, 2012, for commentaries and reviews) in conversational dyads.

The advantages of the beat-tracking measure over traditional measures of speech rate are (a) it is automated and easy to apply to acoustic recordings; (b) it can identify timing regularities despite noisy, irregular patterns that are typical of conversational speech; and (c) it extracts beat rate based on salient moments (i.e., changes of intensity and frequency) of the acoustic speech stream. The beat-tracking algorithm could be viewed as a complimentary measure to traditional measures; it measured beat rate based on regularities between moments of high intensity that generally correspond to speech stress. It remains to be seen whether the algorithm is universally applicable for determining the beat in other languages.

The algorithm's applicability to speech suggests that the same mechanisms responsible for entraining to a musical beat could be involved in the convergence of speech timing (Large & Jones, 1999; Lidji et al., 2011; Wilson & Wilson, 2005). In other words, listeners' neural oscillations may have entrained to regularities in the timing of stressed or accented events. The present results provide some evidence that speech timing convergence might be induced by neural entrainment, but neural measures are required to substantiate this claim. In support of this assertion, evidence from research using EEG measurements suggests that entraining to speech timing regularities in German enhances the prediction of stress patterns and facilitates semantic processing (Rothermich, Schmidt-Kassow, & Kotz, 2012). Moreover, EEG measures during speech perception and production have indicated some synchronization of neural signals between the speaker and the listener (Kuhlen et al., 2012; Stephens, Silbert, & Hasson, 2010) Thus, entrainment to speech timing can occur and might serve a purpose in comprehension.

Behavioral research has shown that the ability to tap in time with a regular metronome, which is often used as a behavioral estimate of neural beat entrainment (Large & Jones, 1999), is related to a range of speech and language abilities; Corriveau and Goswami (2009) found that children with speech and language impairments showed deficits in paced rhythmic tapping compared to those with normal speech, and Tierney and Kraus (2013) showed that reading ability positively correlated with intertap interval regularity. Future studies could measure listeners' cortical responses to speech stimuli to determine if the same neural mechanisms are engaged in speech rate convergence and beat entrainment (as in Nozaradan et al., 2011, 2012).

The present study examined speech rate and beat rate convergence when a confederate purposefully spoke at fast or slow speech rates with participants. In natural conversation, speech rate convergence has been shown to occur (Manson et al., 2013; Street et al., 1983; Webb, 1969, 1972) but has also been elusive under certain circumstances (e.g., Pardo et al., 2010, 2013; Street, 1984). The use of a confederate can sometimes create artifacts that are unlikely to occur in natural conversation, particularly if the participant is suspicious of the confederate (cf. Kuhlen & Brennan, 2012). Participants did not report awareness of the status of the confederate in the present study, and participants' beat rates were altered in response to the confederate's beat rates. Therefore, the beat rate convergence shown in the present study is unlikely to be attributed to participants' suspicion of the confederate. However, the confederate's task of attempting to drive a particular speech rate made the situation less natural. Previous studies that have examined speech rate convergence cannot discount that interlocutors' speech rates may have been similar from the outset (Manson et al., 2013; Street, 1984; Street et al., 1983; Webb, 1969, 1972). By using a confederate, the present study was able to demonstrate that convergence occurred with the same person when the confederate spoke at different speech rates. Thus, the present study avoided confounds stemming from similarities in natural speech rate.

Because the confederate and the majority of the participants in the present study were female, it is possible that the present findings are specific to female dyadic interactions. Pardo (2006) showed that male speakers converged more than female speakers in same-sex conversations in relation to phonetic features, suggesting the current findings would also apply to male speakers. Similarly, Street (1984) demonstrated that speech rates bidirectionally converged for male–female and male–male pairs, but did not converge in either direction for female–female pairs. Therefore, it is possible that male–male or mixed-sex pairs would produce greater speech rate convergence than that observed in the present study.

It is also possible that the use of scripted dialogues may have affected speech rate convergence because the effects of speech planning and comprehension would have been lessened. Although scripts were used for experimental control, it is possible that speech rate could show different effects in conversational speech. For example, Guaïtella (1999) argued that reading aloud induces more regular metric structures in speech (e.g., a beat) as a result of a lack of expression that contributes to the rhythmic organization of spontaneous speech. This assertion, however, appears to be based primarily on readings of nontheatrical scripts, as opposed to the scripts from theatrical plays used in the present study. Because the present studies used scripts from theatrical plays that evoke expression, it is less likely that the beats found in the speech signals were induced by the act of reading aloud.

Beat rate convergence between the confederate and participants increased over the course of the scripted dialogues, and the confederate's rate influenced the participants more than vice versa. Naturally occurring conversational speech might demonstrate bidirectional influences of speakers without such an asymmetry. However, asymmetries may still occur in cases where people are reticent to speak (see Street et al., 1983), when people have disordered speech such as dysarthria (Borrie & Liss, 2014), or when their degree of linguistic competence differs (Beebe & Giles, 1984), for example, between two bilingual speakers speaking in their native or second language. Further work is necessary to determine the impact of individual and social factors on speech convergence within conversations. The beat-tracking algorithm presented here provides a tool to address rate convergence across languages, because it is sensitive only to acoustic cues that influence amplitude.

Conclusions

Speech rate convergence between a confederate and participants was demonstrated using a beat-tracking algorithm that identified regular beats in acoustic recordings of speech, providing evidence that interlocutors entrain to the beat of one another's speech. To the authors' knowledge, this is the first demonstration that a beat tracker can be applied to acoustic speech recordings to measure speech rate. Interpreted through dynamic attending theory (Jones, 2009; Large & Jones, 1999), the results support the proposal that neural oscillations entrain to regularities in the speech signal and affect the rate at which speech is subsequently produced. Furthermore, findings indicated that interlocutors become mutually entrained, as shown by the bidirectional influences of interlocutors' speech rates (Wilson & Wilson, 2005). These findings suggest that speech and music might share similar time-keeping mechanisms and provide an impetus for research into commonalities of the two domains.

ACKNOWLEDGMENTS

This work was supported by Canada Research Chairs awards (to D.T. and C.P.), CRBLM New Initiative Funding (to N.P., D.H.M., D.T., and C.P.), NSERC Grant 261769 (to D.T.), and NSERC Grant 298173 (to C.P.). The authors gratefully acknowledge Fran Spidle, Anna Jamrozic, and Lisa Jewett for assistance with data collection and analysis; the reviewers for their comments and suggestions; and Dan Ellis for technical details on the beat-tracking algorithm.

NOTES

1. The MatLab scripts used to implement the beat tracker can be downloaded from http://www.mcgill.ca/spl/publications.

2. Significant negative correlations were also shown between articulation rate (i.e., syllables per second with pauses removed) and IBI values.

REFERENCES

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*, 351–373.

Beebe, L. M., & Giles, H. (1984). Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language, 46*, 5–32.

Bertinetto, P. M., & Fowler, C. A. (1989). On sensitivity to durational modifications in Italian and English. *Rivista di Linguistica, 1*, 69–94.

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

Bosshardt, H.-G., Sappok, C., Knipschild, M., & Holscher, C. (1997). Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers. *Journal of Psycholinguistic Research, 26*, 425–448.

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between humans and computers. *Journal of Pragmatics, 42*, 2355–2368.

Burgoon, J. K., Stern, L. A., & Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge: Cambridge University Press.

Classé, A. (1939). *The rhythm of English prose.* Oxford: Basil Blackwell.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex, 45*, 119–130.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, *88*, 101–112.

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

Cummins, F. (2012). Looking for rhythm in speech. *Empirical Musicology Review, 7*, 1–2.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26*, 145–171.

Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech, and brain* (pp. 157–166). London: Macmillan.

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 43–95). Waltham, MA: Academic Press.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51–62.

Ellis, D. P. W. (2007a). Beat tracking by dynamic programming. *Journal of New Music Research, 36*, 51–60.

Ellis, D. P. W. (2007b). Music audio tempo estimation and beat tracking. *Dan Ellis: Research*, Retrieved from http://labrosa.ee.columbia.edu/projects/beattrack/

Giles, H. (Ed.) (1977). *Language, ethnicity, and intergroup relations* (pp. 1–370). London: Academic Press.

Gill, S. P. (2012). Rhythmic synchrony and mediated interaction: Towards a framework of rhythm in embodied interaction. *AI & Society, 27*, 111–127.

Grosjean, F., & Lane, H. (1976). How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception & Performance, 2*, 538–543.

Guaïtella, I. (1999). Rhythm in speech: What rhythmic organizations reveal about cognitive processes in spontaneous speech production versus reading aloud. *Journal of Pragmatics, 31*, 509–523.

Huber, J. E. (2008). Effects of utterance length and vocal loudness on speech breathing in older adults. *Respiratory Physiology & Neurobiology, 164*, 323–330.

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General, 141*, 54–75.

Jassem, W., Hill, D. R., & Witten, I. H. (1984). Isochrony in English speech: Its statistical validity and linguistic relevance. *Intonation, Accent and Rhythm*, *8*, 203–225.

Jones, M. R. (2009). Musical time. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The handbook of music psychology* (pp. 81–92). New York: Oxford University Press.

Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, *1–2*, 21–35.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory & Language, 40*, 153–194.

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America, 112*, 2165–2172.

Kuhlen, A. K., Allefeld, C., & Haynes, J. D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, *6*, 266.

Kuhlen, A. K., & Brennan, S. E. (2012). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review, 20*, 54–72.

Large, E. W., Fink, P., & Kelso, S. J. (2002). Tracking simple and complex sequences. *Psychological Research, 66*, 3–17.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*, 119.

Lehiste, I. (1977). Isochrony revisited. *Journal of Phonetics*, *5*, 253–263.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin & Review, 18*, 1035–1041.

London, J. (2012). *Hearing in time* (pp. 3–7). Oxford: Oxford University Press.

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

Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics, 30*, 247–256.

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research, 44*, 128–143.

McKinney, M. F., Moelants, D., Davies, M. E. P., & Klapuri, A. (2007). Evaluation of audio beat tracking and music tempo extraction algorithms. *Journal of New Music Research, 36*, 1–16.

Miller, A. (1949). *Death of a salesman.* New York: Viking Press.

Miller, J. L., & Grosjean, F. (1981). How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 208–215.

Miller, J. L., Grosjean, F., & Lomato, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica, 41*, 215–225.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*, 405–408.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality & Social Psychology, 32*, 790–804.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience, 31*, 10234–10240.

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *Journal of Neuroscience, 32*, 17572–17581.

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

Pardo, J. S., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. (2013). The influence of role-switching on phonetic convergence in conversation. *Discourse Processes, 50*, 276–300.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics, 72*, 2254–2264.

Patel, A. D. (2008). *Music, language, and the brain* (pp. 96–154). New York: Oxford University Press.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*, 411–440.

Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society*, *Series B, 11*, 68–84.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America, 105*, 512–521.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26*, 867–891.

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia, 50*, 232–244.

Schmidt, R. C., Richardson, M. J., Arsenault, C., & Galantucci, B. (2007). Visual tracking and entrainment to an environmental rhythm. *Journal of Experimental Psychology: Human Perception & Performance, 33*, 860–870.

Shen, Y., & Peterson, G. G. (1962). Isochronism in English. *Studies in Linguistics, Occasional Papers*, *9*, 1–36.

Sluijter, A. M., Van Heuven, V. J., & Pacilly, J. J. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America, 101*, 503–513.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Science, 107*, 14425–14430.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). The Mel scale equates the magnitude of perceived differences in pitch at different frequencies. *Journal of the Acoustical Society of America, 8*, 185–190.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research, 11*, 139–169.

Street Jr., R. L., Street, N. J., & Van Kleek, A. (1983). Speech convergence among talkative and reticent three-year-olds. *Language Sciences, 5*, 79–96.

Tierney, A. T., & Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain & Language, 124*, 225–231.

Tukey, J. W. (1958). Bias and confidence in not quite large samples (abstract). *Annals of Mathematical Statistics, 29*, 614.

Villing, R., Ward, T., & Timoney, J. (2003, July 1–2). *P-Centre extraction from speech: The need for a more reliable measure.* Paper presented at the Irish Signals & Systems Conference (ISSC 2003), Limerick, Ireland.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92*, 723–735.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behavior. *Language & Speech, 12*, 54–67.

Webb, J. T. (1972). Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. In B. Pope & A. W. Siegman (Eds.), *Studies in dyadic communication* (pp. 115–133). New York: Pergamon Press.

Wilde, O. (1908). *Collected works of Oscar Wilde.* Metheun, MA: Riverside Press.

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