Speech Rhythm Convergence as a Social Coalition Signal

Leona Polyanskaya[1] , Arthur G. Samuel[1,2,3], and Mikhail Ordin[1,2]

Abstract
Patterns of nonverbal and verbal behavior of interlocutors become more similar as communication progresses. Rhythm entrainment promotes prosocial behavior and signals social bonding and cooperation. Yet, it is unknown if the convergence of rhythm in human speech is perceived and is used to make pragmatic inferences regarding the cooperative urge of the interactors. We conducted two experiments to answer this question. For analytical purposes, we separate pulse (recurring acoustic events) and meter (hierarchical structuring of pulses based on their relative salience). We asked the listeners to make judgments on the hostile or collaborative attitude of interacting agents who exhibit different or similar pulse (Experiment 1) or meter (Experiment 2). The results suggest that rhythm convergence can be a marker of social cooperation at the level of pulse, but not at the level of meter. The mapping of rhythmic convergence onto social affiliation or opposition is important at the early stages of language acquisition. The evolutionary origin of this faculty is possibly the need to transmit and perceive coalition information in social groups of human ancestors. We suggest that this faculty could promote the emergence of the speech faculty in humans.

Evolutionary adaptations allow humans to perceive auditory input as rhythmic and to coordinate their behavior with the acoustic signal (Fitch, 2009; Lang et al., 2016; Large & Snyder, 2009; McNeill, 1995; Merker, Madison, & Eckerdal, 2009; Patel, 2006; Phillips-Silver & Trainor, 2005; Phillips-Silver & Trainor, 2007; Repp & Penel, 2004; Repp & Su, 2013). Social entrainment is a special case of such coordination, representing the entrainment of behavior, including verbal behavior, to the signal emitted by a different conspecific individual (Phillips-Silver, Aktipis, & Bryant, 2010). In social entrainment, mechanisms of rhythmic cognition and synchronization of the motor output with the input signal are activated by the cues from the social environment and allow coordination of movements and vocalizations, including speech production, and even entrainment of neural oscillations (Bowling, Herbst, & Fitch, 2013; Stephens, Silbert, & Hasson, 2010). Rhythmic entrainment in social settings has been claimed to play an important role in social bonding and to promote prosocial behavior (Haidt, Seder, & Kesebir, 2008; Hove & Rinsen, 2009; Kirschner & Tomasello, 2009; McNeill, 1995; Wiltermuth & Heath, 2009).

During verbal interaction, patterns of nonverbal and verbal behavior of interlocutors become more similar as communication progresses (Bargh, Chen, & Burrow, 1996; Chartrand & Bargh, 1999; Dijksterhuis & Bargh, 2001). Convergence of verbal behavior happens at multiple linguistic levels, from phonetic to lexical and syntactic (Pickering & Garrod, 2004). Vocal convergence in human speech has also been a hot topic (Abel & Babel, 2017; Bable, 2012; Pardo, Urmanche, Wilman, & Wiener, 2017; Pardo et al., 2018; Reichel, Benus, & Mady, 2018). Speech rhythm patterns also become progressively more similar between interacting individuals (Beňuš, 2014; Borrie, Barrett, Willi, & Berisha, 2019; Reichel et al., 2018). This

[1] BCBL—Basque Centre on Cognition, Brain and Language, Donostia, Spain
[2] IKERBASQUE—Basque Foundation for Science, Bilbao, Spain
[3] Department of Psychology, Stony Brook University, Stony Brook, NY, USA

Corresponding Author:
Leona Polyanskaya, BCBL—Basque Centre on Cognition, Brain and Language, Mikeletegi 69, Donostia, Gipuzkoa, Spain.
Email: polyanskaya@bcbl.eu

inter-speaker entrainment is modulated by speakers' gender, social status and conversational role, personality, cognitive demands of the interaction, and other factors. Convergence of interlocutors' rhythmic patterns in motor movements and vocalizations has frequently been mentioned as a reliable correlate of communication success and cooperation level: the more similar the rhythmic patterns of the interlocutors are, the higher the communication success and the level of cooperation are (Auer, Couper-Kuhlen, & Müller, 1999; Beattie, Cutler, & Pearson, 1982; Beňuš, 2014; Couper-Kuhlen & Auer, 1991; Couper-Kuhlen, 1993; Cowley, 1994; Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; Street, 1984; Zimmermann & Richardson, 2016). However, such conclusions have only been made so far in speech production studies that cannot cast light on whether listeners can map the rhythmic convergence in speech onto the social cooperation level or whether the convergence is instead an automatic consequence of the general "cooperative urge" of humans. It remains unknown whether listeners can make pragmatic inferences based on the (dis)similarity of rhythmic patterns in speech. This study investigates whether people map similar speech rhythmic patterns onto the level of cooperation and friendliness between the interacting agents.

The term *rhythm* is used in multiple ways, and different researchers may target different phenomena or characteristics of the acoustic signal when they explore rhythm (Goswami & Leong, 2013; Nolan & Jeon, 2014; Ravignani & Morton, 2017). In the current study, rhythm is understood as the structure that determines how the signal is organized and develops over time (McAuley, 2010). In speech, multiple rhythms can operate at multiple levels. For analytical purposes, we separate pulse and meter (Large & Snyder, 2009). Pulse is the occurrence of salient acoustic events. Pulses are used for beat induction, which is the psychological tendency to perceive pulses as equally distributed in time, that is, isochronous (Ravignani & Madison, 2017). Even if the sequence of events is not isochronous, humans tend to regularize the intervals and perceive the sequence of events in the auditory modality as isochronous (Madison & Merker, 2002; McAuley, 2010; Motz, Erickson, & Hetrick, 2013), within certain limits of the interval jitter (Madison & Merker, 2002). Beyond the jitter limits, the percept of isochrony does not emerge. Thus, listeners perceive a series of regularly reoccurring psychological events in response to the auditory stimulation caused by a continuous acoustic input. In continuous speech, vowel onsets are salient acoustic events that generate recurring physiological responses (Greenberg & Ainsworth, 2004) at the frequency of the syllable rate (Ghitza, Giraud, & Poeppel, 2013). These responses are used to extract the syllable as a distinguishable quasi-regular constituent (Ding, Melloni, Zhang, Tian, & Poeppel, 2016) by entraining neural quasi-periodic oscillations to the acoustic rhythm at the syllabic frequency. Perception of regularity in recurrence of vowel onsets is based on the entrainment of neural-to-acoustic oscillations, and this percept facilitates speech comprehension (Assaneo et al., 2019; Ghitza & Greenberg, 2009).

Speech of two interacting agents can elicit either similar or different streams of psychological events, which can potentially be used to make pragmatic inferences regarding the cooperative level between them. We manipulated the distribution of quasi-isochronous syllables and created acoustic syllabic sequences with two distinguishable types of pulse, spoken with different voices. We paired these syllabic sequences with a short pause between them, to imitate brief dialogs in an "alien language"; the paired sequences had either the same or different pulse. We hypothesize that interacting agents producing utterances with similar pulse will be judged as cooperating more than those producing utterances with different pulse. Testing this prediction was the focus of our first experiment.

Meter is the hierarchical organization of salient events in acoustic stream based on their relative salience, that is, grouping the pulses into hierarchical structures (London, 2004). Acoustic perturbations, for example, related to the distribution of relatively more salient sounds in the flow of less salient ones lead to different groupings of the repeated sounds into patterns (Hay & Diehl, 2007). Moreover, meter can even be mentally represented, that is, people can mentally assign different perceptual salience to some of the sounds in the sequence of acoustically identical sounds and group them into patterns based on the prominence levels (Kunert, Willems, Casasanto, Patel, & Hagoort, 2015; Langus, Mehler, & Nespor, 2017, for speech; Patel, 2003b; Patel & Daniele, 2003, for structural similarities and mutual influences between musical and speech rhythms; Palmer & Krumhansl, 1990, for music; Patel, 2003a, for the overlap in the neural substrates and pathways underlying rhythm processing in speech and music). In speech, differences in the distribution of stressed syllables (or acoustic correlates of stress: duration, amplitude, and pitch) may result in different organizations of the syllables into words (Hawthorne, Järvikivi, & Tucker, 2018). For example, in the sequence of syllables [hæ] [pi] [tsə], the syllable [pi] may get grouped either into the word *happy* or *pizza*, depending on the relative distribution of duration and pitch of these three syllables. As every syllable in the speech stream represents a pulse, different distributions of stressed syllables may result in different ways to structure the pulses, that is, to differences in meter.

In the second experiment, we manipulated acoustic prominences, leading to the perception of some syllables as stronger, that is, stressed, relative to the other syllables, that is, unstressed syllables, in the sequence. We manipulated the distribution of stressed syllables to induce the perception of differences in meter, that is, two different ways to group the syllables into metrical structures based on their relative acoustic salience. We hypothesize that meter similarity will result in similar grouping of pulses in utterances produced by interacting agents. If the meter in utterances of interlocutors is the same, listeners will perceive the interacting agents as more cooperative, if pragmatic inferences based on rhythmic synchronization are indeed made. Conversely, if the acoustics of the signals emitted by the interlocutors lead to different groupings of pulses (syllables), then listeners might perceive them as hostile to each other. Our second experiment was focused on

Table 1. Schematic Representation of Sequences for Experiment 1.

Type A pulse	Type B pulse
XX-XXXXX-XX-XXXXX-XX-XXXXX	XXXX-XXX-XXXX-XXX-XXXX-XXX

Note. Each x stands for a syllable, - stands for a pause.

identifying the role of meter synchronization on the pragmatic inferences drawn from speech produced by interacting agents.

Experiment 1: Pulse Synchrony

Method

Participants: Experiment 1

Twenty-six Spanish-Basque native speakers without speech or hearing problems were recruited. The participants were Spanish-dominant bilinguals (age range: 18–30), fully functional in both languages, living in a bilingual environment. This linguistic profile was chosen because it fits the most represented category of residents in the province of Gipuzkoa, the Basque Country, where the experiments were carried out.

Material: Experiment 1

We used four consonants—[s], [m], [n], [l]—and five vowels—[a], [o], [u], [e], [i]—to construct 20 possible consonant–vowel syllables. We added the syllable [fa] to the inventory. The longest stimulus was 21 syllables in length, and we did not want any syllable to be repeated within a stimulus. Thus, we added the syllable [fa] to the inventory to make the inventory size 21 syllables in total. Multi-Band Resynthesis OverLap Add (MBROLA) algorithm (Dutoit & Leich, 1993) speech synthesis software, with Spanish diphone data sets, was used to synthesize the syllables, each in two different voices. Vowel durations were set to 250 ± 15 ms, and consonant durations were set to 100 ms. We prepared ninety 21-syllable sequences with Type A pulse and 90 sequences with Type B pulse (Table 1). Each syllable was used once in each sequence, and the order was unique in each sequence. The duration of pauses within each sequence was set to 300 ms. A declination intonational contour was imposed on each sequence: An utterance-initial high tone rose from 250 Hz to 300 Hz, with the peak aligned on the middle of the second vowel. Then, there was a gradual decline from 300 Hz to 160 Hz over the whole sequence, from the middle of the second vowel to the middle of the penultimate vowel. Finally, there was an utterance-final tone that fell from 160 Hz to 120 Hz from the middle of the penultimate vowel to the end of the final vowel.

Procedure: Experiment 1

The sequences synthesized with different voices were paired with a 1-s pause between them. We paired sequences either with matching pulse (30 stimuli pairing Type A with Type A pulse and 30 stimuli pairing Type B with Type B pulse) or different pulse (30 stimuli pairing Type A vs. Type B pulse and 30 stimuli pairing Type B vs. Type A pulse). Each sequence was used only once, in one stimulus. Participants were told they were going to hear short conversations between two aliens (one alien saying something and the other responding). For each conversation, they were to indicate, on an 8-point scale, whether the aliens are getting along with each other (cooperating) or having a dispute (hostile to each other). The response buttons on the screen were separately grouped into 1 (*definitely hostile*) to 4, and from 5 to 8 (*definitely cooperating*). A short training session (4 stimuli not used in the main study) was run prior to the main experiment to make sure the participants were familiar with the procedure and experimental interface and to establish a comfortable volume level. The experiment was conducted in a soundproof cabin.

We expected the listeners to think that collaborative interacting agents would produce rhythmically similar utterances (due to rhythmic synchronization) and mutually hostile agents would produce rhythmically contrastive utterances (due to the absence of rhythmic synchronization).

We expected that two properties of each participant might affect performance, intelligence, and empathy (the capacity to predict and to respond to the behavior of interacting agents by inferring their mental states). To assess empathy, we asked the participants to fill in two Cambridge Personality Questionnaires (in Spanish) developed at the School of Clinical Medicine, the Department of Psychiatry, University of Cambridge (www.autismresearchcentre.com). The first questionnaire was designed to measure the empathy quotient (EQ, on a scale from 0 to 80); the second one measured the systemizing quotient (SQ, on a scale from 0 to 150). SQ measures the ability to predict the behavior of deterministic systems by inferring the deterministic rules based on an analysis of the systems' input–output relations (Kidron, Kaganovskiy, & Baron-Cohen, 2018). We converted the individual scores for both measures into percentages and calculated the ratio SQ/EQ to estimate individual differences in systemizing drive versus empathizing drive (a lower ratio reflects a stronger emphasizing drive relative to systemizing drive). The stronger the drive, the more it is employed in everyday situations (Baron-Cohen & Wheelwright, 2004). The strength of these two drives relies on the neural architecture (see Kidron et al., 2018, for review), and a stronger systemizing relative to empathizing drive is associated with increased gray matter volume and higher neural activity in certain brain area (Lai et al., 2012), it is also correlated with the level of prenatal testosterone (Auyeung, Lombardo, & Baron-Cohen, 2013; Chapman et al., 2006). These factors suggest that the SQ/EQ ratio is physiologically determined and is not primarily dependent on the ongoing situation and cannot be changed at will when the task requires.

Intelligence (IQ) may affect how well people understand the task and infer that the experimenter wants them to use the rhythmicity to do the task. The participants in the database from which we recruited our sample had already taken the Kaufman Brief Intelligence Test, 2ns Edition (KBIT2)

(Kaufman & Kaufman, 2004). The test is grounded on the fluid reasoning and visual processing theory by Flanagan, McGrew, and Ortiz (2000) for measuring nonverbal IQ scores. High validity of the KBIT intelligence test was reported by Scattone, Raggio, and May (2012).

As we are interested in a fundamental property of cognition, we wanted to control for any individual differences in IQ and the SQ/EQ ratio that might influence the ability to make pragmatic inferences based on rhythmic synchronization. Therefore, we regressed out these variables as covariates in our statistical models.

Results and Discussion: Experiment 1

To explore the effect of rhythmic similarity at the level of pulse on the listeners' perception of cooperation/hostility, we performed an analysis of covariance (ANCOVA) with rhythmic similarity (present vs. absent) as a within-subject factor, SQ/EQ and IQ measures as covariates, and the cooperation-hostility rating as a dependent variable. The results showed that, controlling for the effect of the individual differences in systemizing versus empathizing drive and for IQ, the effect of rhythmic similarity was significant, $F(1, 22) = 11.327$, $p = .003$, $\eta_p^2 = .34$. Cooperation ratings were higher on the stimuli when pulse was the same in paired syllable sequences ($M = 4.8$) compared to when paired sequences exhibited different pulse ($M = 4.7$). The effect of SQ/EQ, $F(1, 22) = .294$, $p = .593$, $\eta_p^2 = .013$, and the effect of IQ, $F(1, 22) = .104$, $p = .75$, $\eta_p^2 = .005$, were not significant. However, both covariates significantly interacted with the presence of synchronization, $F(1, 22) = 6.08$, $p = .022$, $\eta_p^2 = .217$ (for SQ/EQ, Figure 1A) and $F(1, 22) = 7.98$, $p = .01$, $\eta_p^2 = .266$ (for IQ, Figure 1B). The difference in the average cooperation rating for the stimuli with and without rhythmic synchronization was smaller when the covariates' values were larger. That is, the difference in ratings assigned to pairs with similar versus different rhythms (i.e., a measure of sensitivity) was negatively correlated with the SQ/EQ ratio ($r = -.397$, $p = .04$, Figure 2A) and nonverbal IQ ($r = -.472$, $p = .014$, Figure 2B). SQ/EQ ratios and nonverbal IQ scores were not mutually correlated ($r = .004$, $p = .975$) and can be considered statistically independent. The data suggest that the listeners use pulse similarity in the utterances spoken by interacting agents for making pragmatic inferences regarding their mutual cooperation or hostility. However, their judgments are further modulated by the relative strength of systemizing relative to empathizing cognitive style and by nonverbal intelligence. Individuals with stronger EQ relative to SQ make a stronger connection between rhythmic synchronization and cooperation/friendliness. Interestingly, individuals with higher logical IQ scores make weaker connections between rhythmic similarity in the speech of interacting agents and their cooperation/friendliness. These results suggest that the faculty to map rhythmic synchronization in speech on interpersonal affiliation is probably not under conscious control.
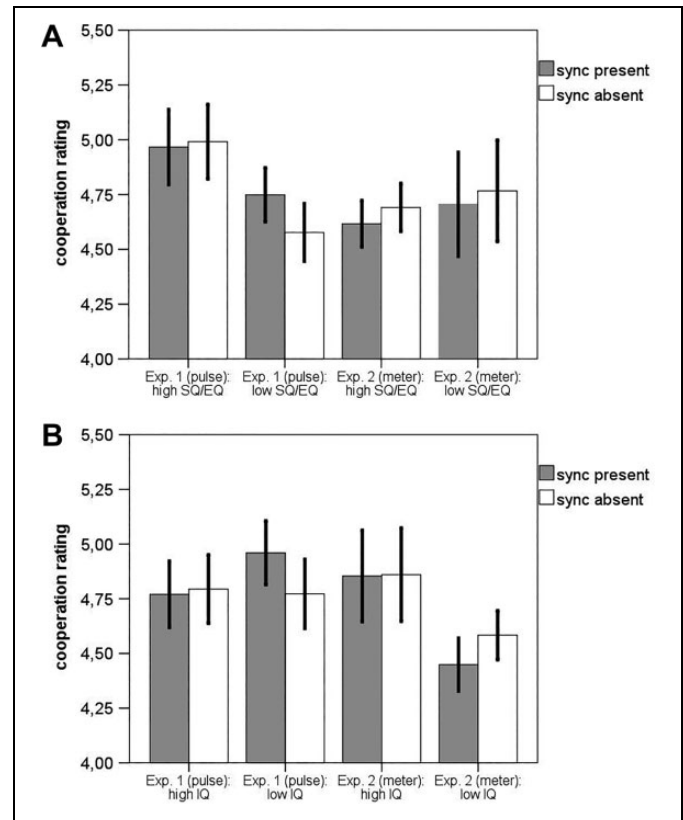

Figure 1. (A) Cooperation ratings based on the median split of the systemizing quotient/empathy quotient (SQ/EQ) ratio values for individual participants. Gray columns display mean cooperation ratings averaged across participants for the trial pairs, in which both interacting agents exhibit the same pulse in utterances, and white columns display mean cooperation ratings averaged across participants for the trial pairs in which interacting agents exhibit different pulse. Error bars (uncorrected for the within-subject design) stand for ±2*SE* around the mean. The data showed that participants with higher SQ relative to EQ values are less sensitive to pulse synchronization as a signal of cooperation. However, this trend is not evident for the pairs of interacting agents with similar versus different meter. (B) Cooperation ratings based on the median split of the nonverbal IQ scores for individual participants. Gray columns display mean cooperation ratings averaged across participants for the trial pairs, in which both interacting agents exhibit the same pulse in utterances, and white columns display mean cooperation ratings averaged across participants for the trial pairs, in which interacting agents exhibit different pulse. Error bars (uncorrected for within-subject design) stand for ±2*SE* around the mean. The data showed that participants with lower nonverbal IQ are more sensitive to pulse synchronization as a signal of cooperation. However, this trend is not evident for the pairs of interacting agents with similar versus different meter.

Experiment 2: Meter Synchrony

Method

Participants: Experiment 2

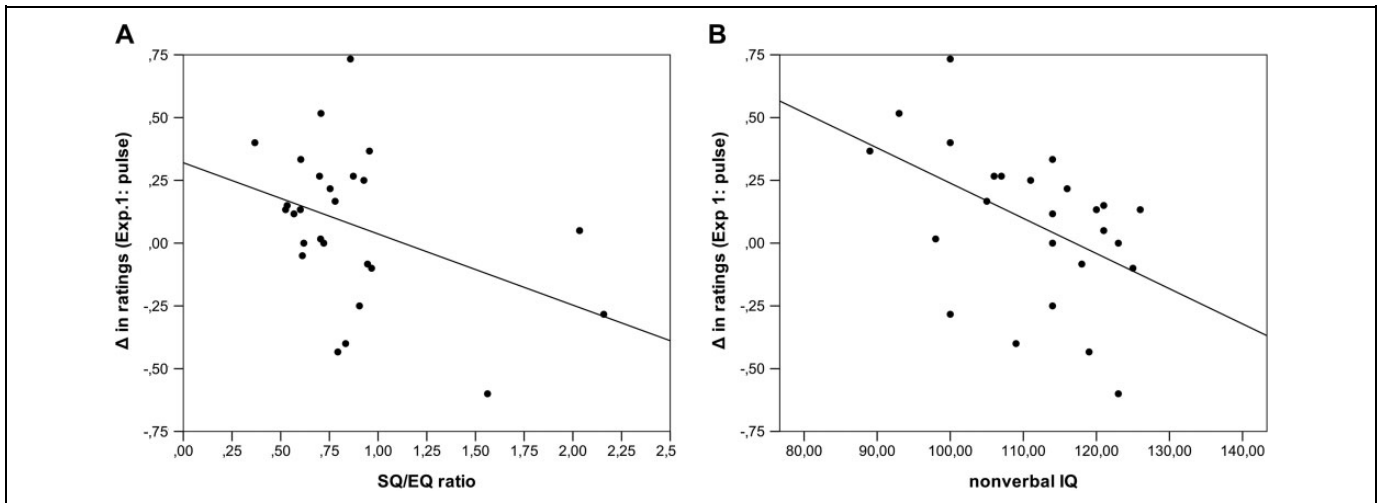Twenty-six participants with the same profile as in the previous experiment were recruited.

Figure 2. (A) Correlations between systemizing quotient/empathy quotient (SQ/EQ) ratio scores and the difference in cooperation ratings assigned to the pairs with similar versus different pulse. The figure shows that higher SQ relative to EQ scores are correlated with lower (or even reverse) differences in cooperation ratings. (B) Correlations between nonverbal IQ scores and the difference in cooperation ratings assigned to the pairs with similar versus different pulse. The figure shows that higher IQ scores are correlated with lower (or even reverse) differences in cooperation ratings.

Table 2. Schematic Representation of Sequences for Experiment 2.

Type A meter	Type B meter
Xxx-Xxx-Xxx-Xxx	xXx-xXx-xXx-xXx

Note. Each X stands for a stressed (strong) syllable, each x stands for unstressed (weak) syllable, - stands for a pause.

Material: Experiment 2

Four consonants ([s], [m], [l],[f]) and 3 vowels ([a], [u], [o]) were used to synthesize 12 possible syllables. We synthesized each syllable in stressed and unstressed versions. The syllables were synthesized with consonantal durations of 100 ms. Vowel durations in stressed, that is, strong syllables, were 300 ms $\pm$ 15 ms, and in unstressed, that is, weak syllables, were 200 ms $\pm$ 15 ms. We synthesized, in two different voices (male, native Spanish phonemes, ES1 and ES3 MBROLA diphone databases were used), 240 syllabic sequences, each composed of 12 syllables. The sequences consisted of four 3-syllable metrical groups, either strong-weak-weak syllables (Meter A) or weak-strong-weak (Meter B) syllables. The groups were separated by a 300 ms pause (Table 2). Three-syllable words in Spanish usually bear prominence on penultimate syllables, while in the regional variety of Basque spoken in San Sebastian, as well as in Batua (the standard variety), the location of word stress is flexible and often depends on the position of the word in a phrase (Hualde, 1999). Therefore, our participants were familiar with both types of metrical grids in their native languages. Each syllable was used once per sequence. An F0 declination contour similar to that in Experiment 1 was imposed on each sequence. Additionally, each strong syllable was made more prominent by a 15 Hz increase, from the middle of the previous syllable and fell back to baseline by the middle of the following vowel, before the declination trend was resumed. Thus, stressed syllables were marked by lengthening of the vowel and by a pitch accent—modeling the shape of an inverted parabola—peaking in the middle of the stressed vowel. The sequences were paired into test stimuli, with a 1-s pause between the paired sequences. The pause duration was chosen based on Ordin, Polyanskaya, Gomez, and Samuel (2019), who used a 1-s pause between stimuli with either similar or different rhythms in an AX rhythm discrimination experiment. We created 30 stimuli with paired sequences exhibiting Type A meter and 30 stimuli with paired sequences with Type B meter, that is, 60 stimuli with similar meter in both sequences. In addition, we created 60 stimuli with different meter in paired sequences, Type A and Type B; the order of the meter type was counterbalanced.

Procedure: Experiment 2

Identical to Experiment 1.

Results and Discussion: Experiment 2

An ANCOVA with meter similarity (present vs. absent) as a within-subject factor and SQ/EQ and IQ measures as covariates was performed to explore the effect of meter synchronization on the cooperation ratings. We did not find any effect of meter synchronization $F(1, 22) = .294$, $p = .593$, $\eta_p^2 = .013$ or nor any for the covariates, $p = .629$ for SQ/EQ and $p = .146$ for nonverbal IQ on the cooperation rating (Figures 1A and 1B). The difference in cooperation rating assigned to pairs with similar versus different meter was not correlated with SQ/EQ ratios (Figure 3A) or nonverbal IQ (Figure 3B). A Bayesian one-tailed paired *t* test was performed to estimate the support for the hypothesis that cooperation ratings assigned to the stimuli
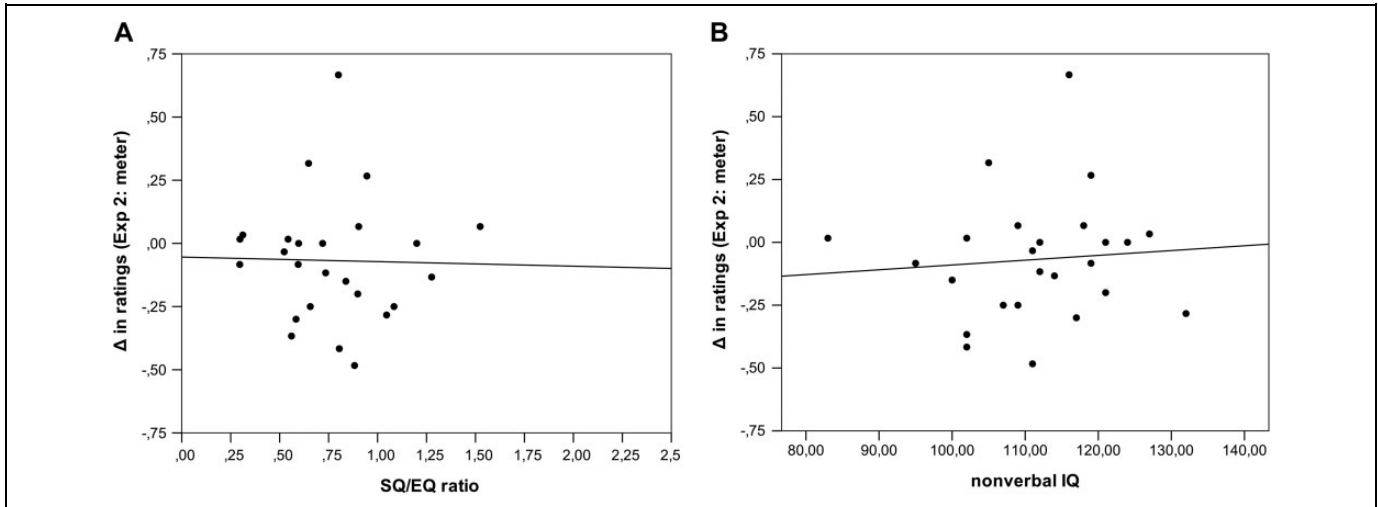
Figure 3. (A) Correlations between SQ/EQ ratio and the differences in cooperation ratings assigned to pairs with similar versus different meters. The figure shows that SQ/EQ scores are not correlated with such differences in cooperation ratings. (B) Correlations between IQ scores and the differences in cooperation ratings assigned to pairs with similar versus different meters. The figure shows that IQ scores are not correlated with such differences in cooperation ratings.

pitting the syllabic sequences with different meter are lower than the ratings assigned to the stimuli exhibiting the same meter in both syllabic sequences. The resulting Bayes factor .098 provided decisive evidence against this experimental hypothesis. The results do not support the hypothesis that listeners use meter similarity in the utterances spoken by interacting agents to make pragmatic inferences regarding their mutual cooperation or hostility.

Finally, for the two experiments, we selected only those stimuli in which the paired sequences had similar pulse or meter and compared, within subject, the number of trials in which the participant responded that the aliens were cooperating (ratings from 5 to 8) with the number of trials in which participants responded that the aliens were hostile (ratings from 1 to 4, Figure 4). For Experiment 1, the data showed that the number of trials in which listeners indicated that the interacting agents were cooperating was higher than the number of trials in which listeners indicated that the interacting agents were hostile, $t(25) = 2.502$, $p = .019$, two-tailed paired test. For Experiment 2, the data did not show a significant difference between the number of trials in which listeners indicated that the interacting agents were cooperating versus hostile, $t(24) = .663$, $p = .513$, two-tailed paired t test. The results confirm that interacting agents producing the utterances with similar pulse are more likely to be judged as cooperating than hostile. However, no evidence that meter similarity is used as a cooperation signal was found.

Discussion

Our study confirms a link between *perception* of social cooperation between individuals and the similarity of rhythmic patterns in their utterances. Listeners map the degree of pulse
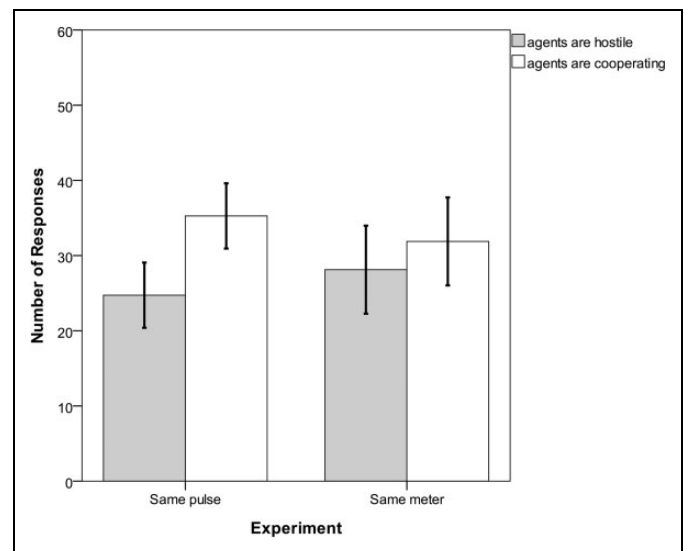

Figure 4. The mean number of trials (averaged across participants) exhibiting pulse (Experiment 1) or meter (Experiment 2) similarity judged as conversations between cooperating and as hostile agents. Error bars $\pm 2SE$ around the mean.

similarity in speech rhythm of two interlocutors onto the degree friendliness and social bonding between them. Conversely, no evidence was found that meter similarity is perceived as a cooperation signal. It should be noted that our data do not resolve whether it is the presence of interpersonal synchronization leads to the perception of cooperation/friendliness versus a difference in rhythmic patterns leading to the perception of hostility. It may be that detecting a similar rhythm in vocalizations of interacting agents results in third-party observers making inferences about their cooperative drive, while rhythmic differences do not lead to making any pragmatic inferences (making the perception of cooperative drive less likely).

Choosing between these alternatives, however, was not part of our experimental design.

The mapping of rhythmic entrainment onto the friendliness of conspecifics may be a "conserved" biological faculty.[1] The referential nature of modern language could have diminished the role of the mapping function tested here because cooperative intention can now be expressed using a Saussurean communication system, instead of (or in addition to) using prosodic means. The traces of this mapping faculty, however, can still be detected today because this faculty has not been selected against. The evolutionary origin of this faculty was possibly the need to transmit coalition information in social groups of human ancestors (Dunbar, 1998; Merker et al., 2009), for example, the rhythmic movements of the lips by primates that usually accompany acts of affiliative behavior (Ghazanfar & Takahashi, 2014a, 2014b; MacNeilage, 1998). The capacity to entrain motor output to acoustic rhythms at the metrical level is rarely exhibited in the animal kingdom, while rhythmic entrainment at the level of pulse is very frequent, at least in mammalian species (Fitch, 2009; Wilson & Cook, 2016). Thus, it is not surprising to observe adoption of a more ancient mechanism to new ecologically relevant sensory input—speech. The faculty of mapping pulse entrainment to cooperation possibly emerged before entrainment at the level of meter developed in some species including the human genus.

For modern humans, the mapping faculty is not essential in everyday interaction due to the referential nature of language. The message is not conveyed by the characteristics of the acoustic signal; rather, the message is conveyed by verbally coding the attributes of a given referent so that a perceiver can easily identify the concept that is referred to (Bowman, 1984, p. 93; Bunce, 1991). As such, the acoustic signal is used to refer to concepts and not to convey the message. However, before the referential system is established, the degree of synchrony can be used by the social partners to develop nonreferential communication and to transmit affection for or discontent with each other, which can be employed at the very early stages of language acquisition. Some strong evidence shows that interactive synchrony is especially important in the early months of life for the development of social cognition in general and speech in particular (Charman, 2005; Tomasello, Carpenter, Call, Behne, & Moll, 2005). When infants are 2 months old, mothers synchronize with and amplify infants' vocalizations, and this behavior is accompanied with sympathy-expressing gestures (smiles, hand gestures) and intermingled with affective vocal expressions (Papousek, 1989). This supportive attunement to infants' vocalizations encourages them to keep on practicing this vocal exchange. About 1 month later, infants begin to recognize the affection expressed by mothers who synchronize their vocalizations with them and start in turn to actively respond in synchronous social interactions (Feldman, 2006; Feldman, Greenbaum, & Yirmiya, 1999). Mothers, in turn, start to attribute intentionality to the attempts of the babies to entrain to the temporal dynamics of social vocal exchanges and create the context for the emergence of intentional vocal exchange expressing affection or discontent (Feldman & Reznick, 1996).

Importantly, speech in live interactions is characterized by coarticulation, vocalic reduction phenomena, by changes in utterance modality (statements vs. questions), so on. These phenomena can affect the salience of rhythmic similarity (Pardo et al., 2018; Reichel et al., 2018), and it remains to be seen whether the demonstrated effect is actually transferrable from laboratory speech to spontaneous interactions. Animals perceive rhythmic cues to make judgments regarding social affiliation (Bergman, 2013; Connor, Smolker, & Bejder, 2006; Ghazanfar & Takahashi, 2014a, 2014b; Ręk & Osiejus, 2010, 2013). Here, we aimed to detect this effect in humans in a situation in which the referential code was not shared by all the parties. We argued that if the referential code is not shared but the effect is still present, then it possibly had evolutionary value at the stage before language and a common referential communicative system emerged. We found that when rhythmic similarity is the only available cue to do an explicitly formulated task, humans can make pragmatic inferences based on speech rhythm in signals produced by interacting agents. It might be that humans have preserved this faculty only because it has not been selected against since the time speech emerged. Should that be true, then, in more ecological situations, the role of speech rhythm in the perception of cooperation between interlocutors might be overshadowed by other signals including referential signals for social affiliation. If, on the other hand, natural selection is still acting on this faculty, then the perception of rhythmic similarity for cooperation judgments is still useful, and we should be able to observe the mapping effect in natural speech.

The importance of mapping the degree of rhythmic similarity onto the degree of interpersonal affection in ontogenetic development of speech suggests that it could also play a role in speech emergence in phylogenetic development of the language faculty in human genus. Knight (2000) argued that speech could have only emerged as a cooperative signaling system. Noble (2000) tested this hypothesis by modeling the emergence of prelinguistic communication. He concluded that a complex and efficient communicative system can emerge and develop only by increasing the fitness of both signaler and receiver. This can only happen in the case of cooperation and thus cannot emerge when there is a conflict of interest, that is, from a costly manipulative signaling system. Oliphant (1996) came to a similar conclusion: His simulations showed that a shared referential communication system could emerge only when both the signaler and the receiver cooperate in making the message transparent. When only the receiver or transmitter is under pressure to convey the message, a shared Saussurean communication system failed to emerge, even though both interacting agents would benefit from it. Rhythmic synchronization can play a role in establishing a common shared communication system by promoting cooperative behavior and social affiliation. The effect of rhythm and rhythmic synchronization in vocal signaling on establishing and developing communicative systems is an interesting
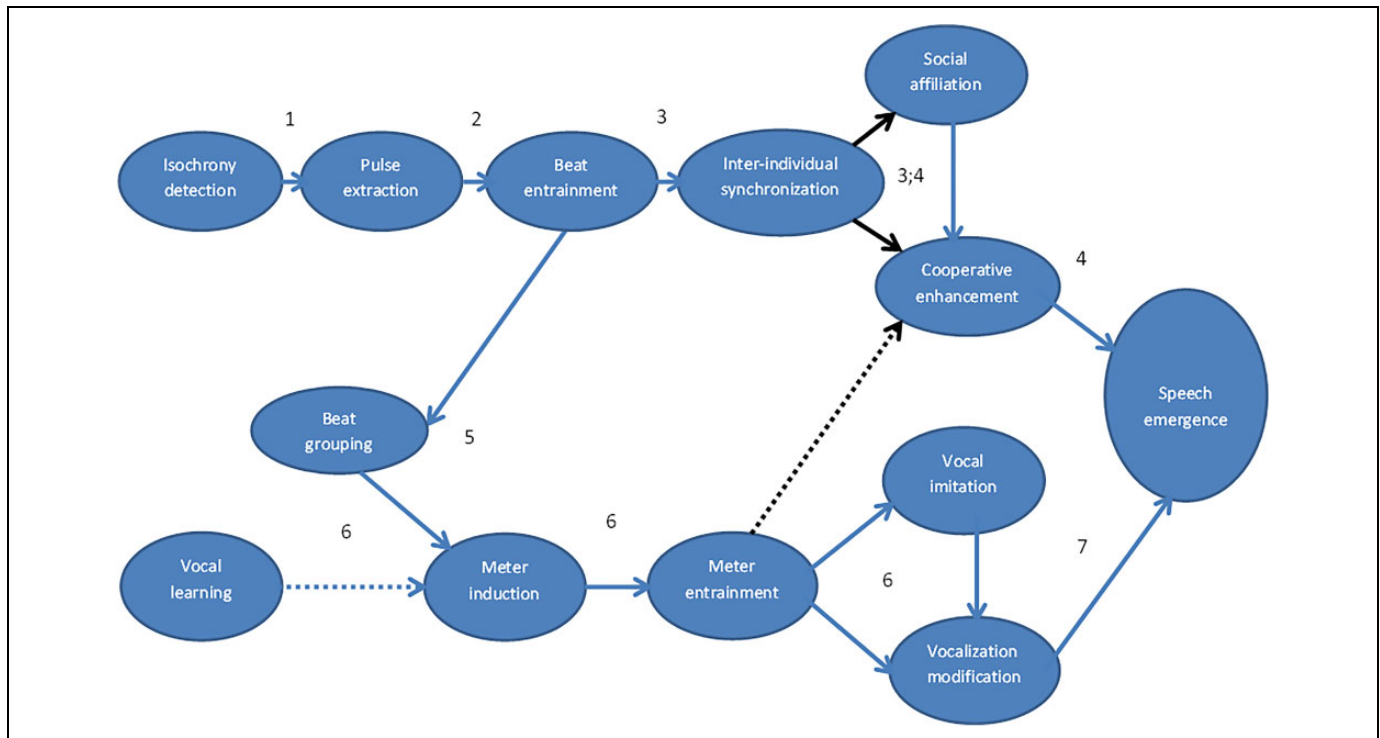
Figure 5. The role of rhythmic cognition in speech emergence from an evolutionary perspective. Dotted lines stand for controversial causal links (i.e., those for which no empirical evidence or inconsistent empirical evidence exists). Black lines show the causal links that are directly tested in the present experiments. Numbers stand for some references that support the corresponding causal links: (1) Ravignani and Madison, 2017, (2) Patel et al., 2009, (3) Koban, Ramamoorthy, and Konvalinka, 2019, (4) Tomasello et al., 2005, (5) Nozaradan, Peretz, Missal, and Mouraux, 2011, (6) Fitch, 2013, (7) MacNeilage, 1998.

research direction that can potentially provide us with insights into the emergence of such complex communication systems as speech. In Figure 5, we present an evolutionary perspective on the role of different components of rhythmic cognition in speech emergence, which can be used for generating testable hypotheses.

Our study confirmed the link between social interactions and speech rhythm. Humans can perceive and synchronize with rhythms in vocalizations emitted by an interlocutor, and outside observers make pragmatic inferences regarding whether the interacting individuals are in mutually hostile or in friendly and cooperating relationships, depending on whether the rhythms in their utterances are similar or different. However, this effect was found only at the level of pulse and not at the level of meter. This suggests that pulse convergence in vocalizations can signal social cooperation. The mapping of vocal rhythm convergence onto social affiliation is important for the development of social cognition and for language acquisition in ontogenesis and probably was an important facilitating factor for speech emergence in phylogenesis.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID ID

Leona Polyanskaya https://orcid.org/0000-0002-3702-786X

Note

1. It was brought to our attention during the peer-review process that the tendency of cooperating individuals to entrain could be learned through experience and not be a biological faculty. We admit that this alternative cannot be ruled out completely. However, we have good reasons to believe that the mapping of interpersonal rhythmic entrainment onto friendliness is not a learned-through-experience faculty. As we discuss later, this faculty is employed by infants as young as 2 months of age, and it is a crucial prerequisite for the development of social intelligence (e.g., Charman, 2005; Feldman, 2006; Tomasello et al., 2005). Besides, if this mapping faculty were learned through experience, we would expect individuals with a higher systemizing relative to empathizing quotient to outperform those with higher empathizing relative to systemizing quotient

because they are better at extracting and systemizing the rules for the use in future situations. We, however, found the reverse trend. More studies are necessary to explore whether, or to what extent, differences in individual experience affect the strength of the mapping between interpersonal rhythm entrainment and cooperative drive/friendliness.

References

Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*, *60*, 479–502.

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, *22*, 627–632.

Auer, P., Couper-Kuhlen, E., & Müller, F. (1999). *Language in time: The rhythm and tempo of spoken interaction*. Oxford, England: Oxford University Press.

Auyeung, B., Lombardo, M., & Baron-Cohen, S. (2013, May). Prenatal and postnatal hormone effects on the human brain and cognition. *Pflügers Archiv—European Journal of Physiology*, *465*, 557–571.

Bable, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*, 177–189.

Bargh, J., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype priming on action. *Journal of Personality and Social Psychology*, *71*, 230–244.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163–175.

Beattie, G. W., Cutler, A., & Pearson, M. (1982). Why is Mrs. Thatcher interrupted so often? *Nature*, *300*, 744–747.

Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognition Computing*, *6*, 802–813.

Bergman, T. (2013). Speech-like vocalized lip-smacking in geladas. *Current Biology*, *23*, R268–R269.

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

Bowling, D., Herbst, C., & Fitch, T. (2013). Social origin of rhythm? Synchrony and temporal regularity in human vocalization. *PLoS One*, *8*, e80402.

Bowman, S. (1984). A review of referential communication skills. *Australian Journal of Human Communication Disorders*, *12*, 92–112.

Bunce, B. H. (1991). Referential communication skills: Guidelines for therapy. *Language, Speech, and Hearing Services in Schools*, *22*, 296–301.

Chapman, E., Baron-Cohen, S., Auyeung, B., Knickmeyer, R., Taylor, K., & Hackett, G. (2006). Fetal testosterone and empathy: Evidence from the empathy quotient (EQ) and the 'reading the mind in the eyes' test. *Social Neuroscience*, *1*, 135–148.

Charman, T. (2005). Why do individuals with autism lack the motivation or capacity to share intentions? *Behavioral and Brain Sciences*, *28*, 695–696.

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

Connor, R. C., Smolker, R., & Bejder, L. (2006). Synchrony, social behaviour and alliance affiliation in Indian Ocean bottlenose dolphins, Tursiops aduncus. *Animal Behaviour*, *72*, 1371–1378.

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction*. Amsterdam, the Netherlands: John Benjamins.

Couper-Kuhlen, E., & Auer, P. (1991). On the contextualizing function of speech rhythm in conversation: Question–answer sequences. In J. Verschueren (Ed.), *Levels of linguistic adaptation: Selected papers of the international pragmatics conference*, Antwerp, 1987, 1–18. Amsterdam, the Netherlands: John Benjamins.

Cowley, S. J. (1994). Conversational functions of rhythmical patterning: A behavioral perspective. *Language and Communication*, *14*, 353–376.

Dijksterhuis, A., & Bargh, J. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 33, pp. 1–40). San Diego, CA: Academic Press.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*, 158–164.

Dunbar, R. (1998). The social brain hypothesis. *Evolutionary Anthropology*, *6*, 178–190.

Dutoit, T., & Leich, H. (1993). MBR-PSOLA: Text-to-speech synthesis based on an MBE re-synthesis of the segments database. *Speech Communication*, *13*, 435–440.

Feldman, R. (2006). From biological rhythms to social rhythms: Physiological precursors of mother-infant synchrony. *Developmental Psychology*, *42*, 175–188.

Feldman, R., Greenbaum, C. W., & Yirmiya, N. (1999). Mother-infant affect synchrony as an antecedent to the emergence of self-control. *Developmental Psychology*, *35*, 223–231.

Feldman, R., & Reznick, J. S. (1996). Maternal perception of infant intentionality at 4 and 8 months. *Infant Behavior and Development*, *19*, 485–498.

Fitch, T. (2013). Rhythmic cognition in humans and animals: Distinguishing meter and pulse perception. *Frontiers in System Neuroscience*, *7*, 68.

Fitch, W. T. (2009). *The biology and evolution of rhythm: Unraveling a paradox. Language and music as cognitive systems*. Oxford, England: Oxford University Press.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and GF-GC theory: A contemporary approach to interpretation*. Boston, MA: Allyn & Bacon.

Ghazanfar, A., & Takahashi, D. (2014a). Facial expressions and the evolution of the speech rhythm. *Journal of Cognitive Neuroscience*, *26*, 1196–1207.

Ghazanfar, A., & Takahashi, D. (2014b). The evolution of speech: Vision, rhythm, cooperation. *Trends in Cognitive Sciences*, *18*, 543–553.

Ghitza, O., Giraud, A.-L., & Poeppel, D. (2013). Neuronal oscillations and speech perception: Critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience*, *6*, 340.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*, 113–126.

Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives. *Laboratory Phonology*, *4*, 67–92.

Greenberg, S., & Ainsworth, W. (2004). Speech processing in the auditory system: An overview. In S. Greenberg, W. Ainsworth, A. Popper, & R. Fay (Eds.), *Speech processing in the auditory system* (pp. 1–62). New York, NY: Springer-Verlag.

Haidt, J., Seder, J., & Kesebir, S. (2008). Hive psychology, happiness, and public policy. *Journal of Legal Studies*, *37*, 133–156.

Hawthorne, K., Järvikivi, J., & Tucker, B. (2018). Finding word boundaries in Indian english-accented speech. *Journal of Phonetics*, *66*, 145–160.

Hay, J. S., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics*, *69*, 113–122.

Hove, M., & Risen, J. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*, 949–961.

Hualde, J. (1999). Basque accentuation. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 947–993). Berlin, Germany: Mouton de Gruyter.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kidron, R., Kaganovskiy, L., & Baron-Cohen, S. (2018). Empathizing-systemizing cognitive styles: Effects of sex and academic degree. *PLoS One*, *13*, e0194515.

Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, *102*, 299–314.

Knight, C. (2000). The evolution of cooperative communication. In C. Knight, J. R. Hurford, & M. Studdert-Kennedy (Eds.), *The evolutionary emergence of language: Social function and the origin of linguistic form* (pp. 19–26). Cambridge, England: Cambridge University Press.

Koban, L., Ramamoorthy, A., & Konvalinka, I. (2019). Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle. *Social Neuroscience*, *14*, 1–9.

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS One*, *10*, e0141069. doi:10.1371/journal.pone.0141069

Lai, M., Lombardo, M., Chakrabarti, B., Ecker, C., Sadek, S., & Baron-Cohen, S. (2012). Individual differences in brain structure underpin empathizing–systemizing cognitive styles in male adults. *Neuroimage*, *61*, 1347–1354.

Lang, M., Shaw, D., Reddish, P., Wallot, S., Mitkidis, P., & Xygalatas, D. (2016). Lost in the rhythm: Effects of rhythm on subsequent interpersonal coordination. *Cognitive Science*, *40*, 1797–1815.

Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. *Neuroscience and Behavioural Reviews*, *81B*, 158–166.

Large, E., & Snyder, J. (2009). Pusle and meter as neural resonances. *Annals of the New York Academy of Sciences*, *1169*, 46–57.

London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. Oxford, England: Oxford University Press.

MacNeilage, P. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, *21*, 499–511.

Madison, G., & Merker, B. (2002). On the limits of anisochrony in pulse attribution. *Psychological research*, *66*, 201–207.

McAuley, J. D. (2010). Tempo and rhythm. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Springer handbook of auditory research, 36: Music perception*. New York, NY: Springer-Verlag.

McNeill, W. (1995). *Keeping together in time: Dance and drill in human history*. Cambridge, MA: Harvard University Press.

Merker, B., Madison, G., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, *45*, 4–17.

Motz, B. A., Erickson, M. A., & Hetrick, W. P. (2013). To the beat of your own drum: Cortical regularization of non-integer ratio rhythms toward metrical patterns. *Brain and Cognition*, *81*, 329–336.

Noble, J. (2000). Co-operation, competition and the evolution of pre-linguistic communication. In C. Knight, J. R. Hurford, & M. Studdert-Kennedy (Eds.), *The evolutionary emergence of language: Social function and the origin of linguistic form* (pp. 40–61). Cambridge, England: Cambridge University Press.

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Transactions of the Royal Society*, *B*, *369*. doi:10.1098/rstb.2013.0396

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

Oliphant, M. (1996). The dilemma of saussurean communication. *Biosystems*, *37*, 31–38.

Ordin, M., Polyanskaya, L., Gomez, D., & Samuel, A. (2019). The role of native language and the fundamental design of the auditory system in detecting rhythm changes. *Journal of Speech, Language and Hearing Research*, *62*, 835–852.

Palmer, C., & Krumhansl, C. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 728–741.

Papousek, M. (1989). Determinants of responsiveness to infant vocal expression of emotional state. *Infant Behavior and Development*, *12*, 507–524.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talker. *Attention, Perception and Psychophysics*, *79*, 637–659.

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

Patel, A. D. (2003a). Language, music, syntax and the brain. *Nature Neuroscience*, *6*, 674–681.

Patel, A. D. (2003b). Rhythm in language and music: Parallels and differences. *Annals of the New York Academy of Sciences*, *999*, 140–143.

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, *24*, 99–104.

Patel, A. D., & Daniele, J. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*, B35–B45.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, *19*, 827–830.

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

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, *308*, 1430.

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, *105*, 533–546.

Pickering, M., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Bahavioral and Brain Sciences*, *27*, 169–226.

Ravignani, A., & Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology*, *8*, 1820.

Ravignani, A., & Morton, P. (2017). Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *Journal of Language Evolution*, *1*, 4–19.

Reichel, U. D., Beňuš, Š., & Mády, K. (2018). Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication*, *100*, 46–57.

Ręk, P., & Osiejuk, T. S. (2010). Sophistication and simplicity: Conventional communication in a rudimentary system. *Behavioral Ecology*, *21*, 1203–1210.

Ręk, P., & Osiejuk, T. S. (2013). Temporal patterns of broadcast calls in the corncrake encode information arbitrarily. *Behavioral Ecology*, *24*, 547–552.

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*, 403–452.

Repp, B., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, *68*, 252–270.

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

Scattone, D., Raggio, D. J., & May, W. (2012). Brief report: Concurrent validity of the leiter-R and KBIT-2 scales of nonverbal intelligence for children with autism and language impairments. *Journal of Autism and Developmental Disorders*, *42*, 2486–2490.

Stephens, G., Silbert, L., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *PNAS*, *107*, 14425–14430.

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

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–735.

Wilson, M., & Cook, P. F. (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychonomic Bulletin and Review*, *23*, 1647–1659.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, *20*, 1–5.

Zimmermann, J., & Richardson, D. (2016). Verbal synchrony and action dynamics in large groups. *Frontiers in Psychology*, *7*, 2034.