

# Word-by-word entrainment of speech rhythm during joint story building

Tommi Himberg\*, Lotta Hirvenkari, Anne Mandel and Riitta Hari

Brain Research Unit, Department of Neuroscience and Biomedical Engineering, School of Science, Aalto University, Espoo, Finland

## OPEN ACCESS

### Edited by:

Judith Holler,  
Max Planck Institute for  
Psycholinguistics, Netherlands

### Reviewed by:

Michael Richardson,  
University of Cincinnati, USA  
Fred Cummins,  
University College Dublin, Ireland

### \*Correspondence:

Tommi Himberg,  
Department of Neuroscience and  
Biomedical Engineering, School of  
Science, Aalto University,  
Puumiehenkuja 2, FI-02150, Espoo,  
Finland  
tommi.himberg@aalto.fi

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 04 March 2015

**Accepted:** 27 May 2015

**Published:** 12 June 2015

### Citation:

Himberg T, Hirvenkari L, Mandel A and  
Hari R (2015) Word-by-word  
entrainment of speech rhythm during  
joint story building.  
*Front. Psychol.* 6:797.  
doi: 10.3389/fpsyg.2015.00797

Movements and behavior synchronize during social interaction at many levels, often unintentionally. During smooth conversation, for example, participants adapt to each others' speech rates. Here we aimed to find out to which extent speakers adapt their turn-taking rhythms during a story-building game. Nine sex-matched dyads of adults (12 males, 6 females) created two 5-min stories by contributing to them alternately one word at a time. The participants were located in different rooms, with audio connection during one story and audiovisual during the other. They were free to select the topic of the story. Although the participants received no instructions regarding the timing of the story building, their word rhythms were highly entrained ( $\bar{R} = 0.70$ ,  $p < 0.001$ ) even though the rhythms as such were unstable ( $\bar{R} = 0.14$  for pooled data). Such high entrainment in the absence of steady word rhythm occurred in every individual story, independently of whether the subjects were connected via audio-only or audiovisual link. The observed entrainment was of similar strength as typical entrainment in finger-tapping tasks where participants are specifically instructed to synchronize their behavior. Thus, speech seems to spontaneously induce strong entrainment between the conversation partners, likely reflecting automatic alignment of their semantic and syntactic processes.

**Keywords:** turn-taking, entrainment, word rhythm, mutual adaptation, speech, social interaction

## Introduction

During human social interaction, body movements and behavior synchronize at many levels. This interpersonal coordination can be intentional or unintentional, and it can take many shapes. In conversation, participants' utterance length, vocabulary, and information density, as well as body posture and the use of non-verbal gestures often adapt or match (Condon and Ogston, 1967; Kendon, 1970; Giles et al., 1991; Chartrand and Bargh, 1999; Gonzales et al., 2010). Similarly, continuous rhythmic behaviors can entrain, or converge in phase and period (Pikovsky et al., 2001; Clayton et al., 2004). Such an entrainment to a common rhythm can be seen in music and dance, finger tapping, rocking in chairs or gait when walking side by side (Boker et al., 2005; Repp, 2005; Richardson et al., 2007; Nessler and Gilliland, 2009; Himberg and Thompson, 2011). Entrainment has positive affective consequences (Hove and Risen, 2009; Wiltermuth and Heath, 2009), and while foregrounded in music and dance, timing and entrainment also play important roles in verbal and non-verbal communication (Bavelas et al., 1986; Shockley et al., 2003; Cummins, 2009).

During smooth conversation, turn-taking is accurately regulated between the participants, who thereby can avoid overlap of speech and optimize silence between the turns. To time their own contributions correctly, the participants need to be able to predict the end of their partner's turn. Traditionally, turn-taking is said to be governed by a set of linguistic rules (Sacks et al., 1974),

while more contemporary theories have suggested turn-taking to be driven by entrainment of oscillatory processes (Wilson and Wilson, 2005), and to operate at the level of prosody and timing, rather than linguistic units (Cowley, 1998). Turn-taking is often seen as fundamental in human cognition, even as a species-specific, evolutionary adaptation (Sidnell, 2001). The basic mechanisms of turn-taking are thought to be universal, although different languages somewhat vary in the optimal duration of gaps between turns (Stivers et al., 2009).

Interpersonal entrainment is a result of continuous mutual adaptation, as has been demonstrated in simple hand-tapping tasks performed by two persons (Konvalinka et al., 2010) as well as in dance (Himberg and Thompson, 2011). Such mutual adaptation emerges in live dyadic interaction and can be observed already in infants (Malloch and Trevarthen, 2009). For example, when participants read texts together, their verbal outputs are better synchronized when they are in live interaction than when they co-read with recorded speech (Cummins, 2009). Moreover, partners synchronize their finger-tapping better with other humans than with non-responsive computer partners (Himberg, 2014).

Interpersonal coordination in dyads and groups can either occur by matching behaviors, such as gestures, posture, or vocabulary, or as continuous synchronization (Bernieri and Rosenthal, 1991; Dale et al., 2013). Both types of coordination occur in natural conversations, but from an experimental perspective, both have complications. Behavior matching, although commonly observed in many aspects of conversations (e.g. as imitation of the other person's actions, called "chameleon effect" by Chartrand and Bargh, 1999), occurs intermittently, as the interlocutors do not mirror each other's contributions, but rather interact in a complementary fashion (Abney et al., 2014). Also, the time lags of matching are unpredictable, and can be as long as minutes (Louwerse et al., 2012). Continuous synchronization also occurs during natural conversations, for example, the body sways of the interlocutors synchronize. However, these movements are so small that measuring them requires special sensors, and even then the signal is noisy and the data analysis is complicated (Shockley et al., 2003). To overcome these complications, we used a word game where turns change predictably and often enough, and thus we could measure interpersonal coordination from the speech signals.

Our aim was to experiment on interpersonal coordination using a linguistic task, to contrast with the cognitively less challenging finger-tapping tasks that are the traditional approaches to studying intentional synchrony (Repp, 2005). We aimed at a task that would feel natural and be easy to explain to the subjects and would allow us to measure interpersonal synchronization directly from the speech signals, rather than relying on changes in secondary, oscillatory movements, such as swinging a pendulum or rocking in a chair (Richardson et al., 2005, 2007). Unlike Reich et al. (2014) who looked at pitch synchrony between therapists and clients, we were interested in word timing. We thus asked pairs of participants to create stories word by word, each contributing one word at a time. Since turn-taking occurred after every word, we were able to study word timing in a relatively controlled situation. As Finnish is a highly

inflected language, each turn consisted of a meaningful word, rather than a preposition, article etc. that do not exist in Finnish (see Supplementary Information 1). Our participants were seated in separate rooms and connected via either an audiovisual link ("video call") or audio-only link ("telephone call"), allowing us to analyze the relative contributions of auditory and visual cues to speech-rhythm entrainment. The terminology and criteria regarding synchronization and entrainment vary largely in the literature (for a review, see Himberg, 2014, pp. 21–35), but in the present study, by word-rhythm entrainment, we refer to phase-locking of the temporal sequences of word onset times of the two participants.

## Methods

### Participants, Apparatus, Materials

We studied 18 healthy adults (12 males, 6 females; aged 21–43 years, mean  $\pm$  SD 27.1  $\pm$  0.6 years), all native Finnish speakers, forming 9 sex-matched pairs. After the course of the study had been explained to the subjects, they gave their written informed consent. The study had prior approval by the Ethics Committee of the Hospital District of Helsinki and Uusimaa.

The data were collected during a two-person magnetoencephalography (MEG) experiment, using a MEG2MEG setup (Baess et al., 2012) but only the behavioral results will be reported here. Participants were seated in separate rooms and, depending on the task condition, they had either an audio-only connection (microphones and headphones), or an audiovisual connection where they could also see a video feed of the other participant in natural size on a projection screen positioned 1 m in front of them.

In our custom-made internet-based communication system, the one-way latency is 50  $\pm$  2 ms for audio signal and 130  $\pm$  12 ms for video (Zhdanov et al., in press). In a pilot dyad, the participants reported they did not notice any lags in either audio or video transmission, and they rated the feeling of presence of their partner at 9 on a 10-point scale. Our participants also reported not to have detected the 80-ms asynchrony between the audio and video inputs during normal conversation that was also included in the setup. This feeling of real-life-like presence of the other person is understandable because the audio and video latencies of our system were well under the limits for smooth conversation (100 ms for audio, 500 ms for video; Jansen and Bulterman, 2013), and even under the limits for more delay-sensitive tasks (60 and 140 ms; Kurita et al., 1994). The asynchrony between the audio and video inputs was within the 130-ms integration window within which auditory and visual speech inputs are considered synchronous, when the auditory input precedes the visual one (Dixon and Spitz, 1980; Vroomen and Stekelenburg, 2011). We therefore considered the transmission latencies of our setup to be negligible for our task, where inter-word intervals were over 2 s.

### Procedure

Participants were instructed to construct a story, contributing one word at a time in alternating turns. They were free to select the topic of the story, and no instructions were given regarding

the rhythm or the timing of the words. The experimenter indicated which participant was supposed to start. The stories were about 5 min in duration. Each dyad constructed two stories, one in which they had only an audio connection, and another where they also could see each other on screen. The order of conditions was counterbalanced across dyads. Because of time constraints, two of the nine dyads completed the task in only one of the two conditions, leaving 16 stories to be analyzed.

## Analysis

We aimed to quantify the rhythm of speech and the interdependence of word timing both for each single individual and between the participants of a dyad. A total of seven instances of coughing, laughing and interruptions due to not hearing the word were removed from the data.

In speech, the stream of stressed syllables generates the word rhythm (Vos et al., 1995; Scott, 1998). In Finnish, word stress occurs on the first syllable of the word (Iivonen, 1998), and therefore we opted to use word onsets as the basis of our word-rhythm analysis. Word onset and offset times were defined in Matlab 7 (MathWorks) from the 48-kHz audio files as the moments where the sound envelope exceeded the level of background noise during silence. After the detection of the onset and offset times, each sound was labeled manually as a word or a non-word and then transcribed. If the actual word was preceded by an interjection (participant saying e.g. “umm... fishing”), the beginning of the interjection was selected as the onset time, as in many cases the interjection and the word were inseparably merged.

Four different time series were extracted from the word-onset and -offset data (Figure 1A): inter-turn intervals (ITIs; times between consecutive word onsets for one speaker), inter-word intervals (IWIs; times between successive word onsets in the joint stream), word durations (DURs), and gap durations (GAPs).

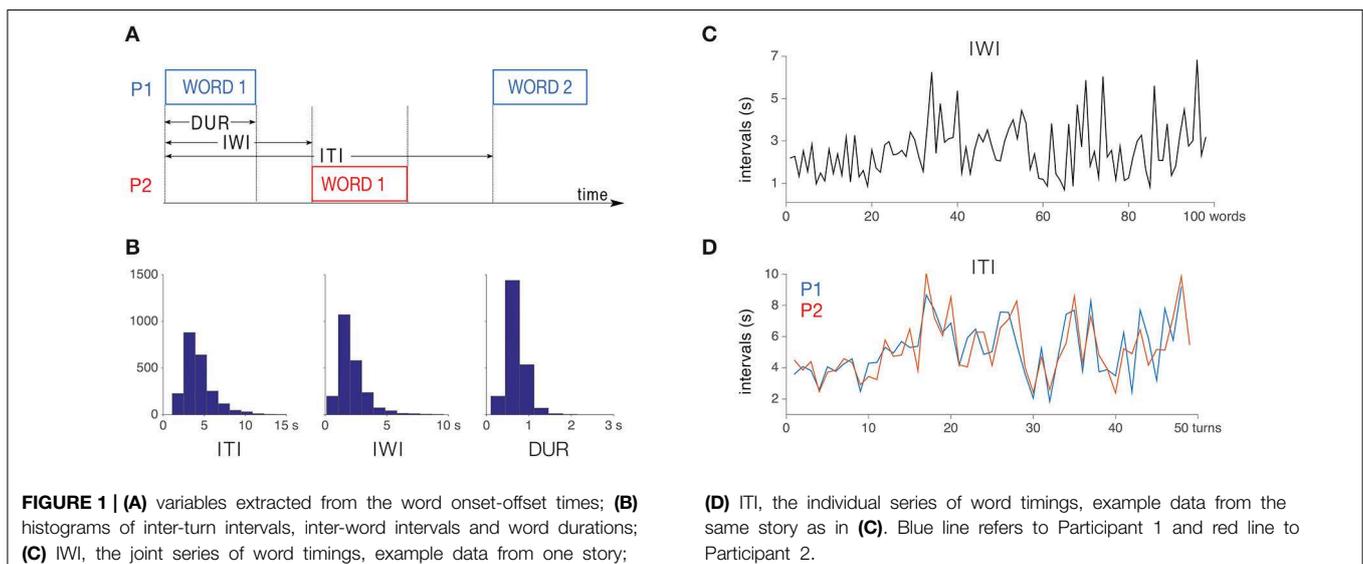
The IWI and ITI time series were converted to phase values, and the concentrations of the resulting circular

distributions were used as stability and entrainment measures (see Supplementary Information 2). The stability measure represented the “steadiness” of the consecutive IWIs or ITIs, quantifying how similar each time interval was in relation to the previous one: equally long intervals yielded a phase value of zero, while deviations yielded non-zero values ranging from 1 to 359°.

The entrainment measure, on the other hand, reflected the consistency of the interrelationship between the ITIs of the two participants. It was calculated by measuring where, within one participant’s ITIs, the other participant’s word onsets occurred. If Participant 1 uttered a word at exactly half way the ITI of Participant 2, the phase value was 180°, with deviations from this anti-phase state ranging theoretically from 0 to 359°. In practice, however, the possible range of relative angles was somewhat narrower (we observed it to range from 14 to 326°), because the participants needed to avoid overlaps (zero relative phase would mean that both participants would start their words simultaneously).

For both stability and entrainment, circular distribution measure  $\bar{R}$  (Fisher, 1993, p. 32) and mean angle  $\theta$  were calculated for each trial, as well as for all the data of the experiment.  $\bar{R}$  ranges from 0 (no stability or no entrainment) to 1 (perfect stability or perfect entrainment), and it has previously been used in quantifying individual timing stability and especially synchronicity and entrainment in dyadic and group timing (Himberg, 2006, 2014; Rankin et al., 2009; Lucas et al., 2011).

To statistically evaluate whether the word rhythms in trials were stable and/or entrained, we conducted *V*-tests and Kuiper two-sample tests to see if the observed distributions statistically differed from uniform distributions (Fisher, 1993; Jammalamadaka and Sengupta, 2001). For the entrainment measure, we compared the observed distribution with a uniform distribution from 14 to 326°, corresponding to the range of phase angles that was observed in the study.



## Results

### General

In the 16 stories by 9 different dyads, a total of 2261 words were uttered, on average 141.3 words per story, or 70.7 (range 43–110) words per participant per story. **Figure 1B** shows the histograms for word durations, IWIs, and ITIs. The word durations were on average (mean  $\pm$  SD)  $0.69 \pm 0.23$  s, IWIs were  $2.14 \pm 1.15$  s, and ITIs were  $4.29 \pm 1.80$  s.

Participants produced the words in a normal tempo with a mean rate of 3.3 syllables/s, which is comparable to that of normal spoken Finnish (Toivola et al., 2009). As expected, due to the nature of the task, the gaps between words (on average 1.45 s) were longer than in normal, continuous speech (0.5 s, Toivola et al., 2009). The sentences that the participants constructed together were syntactically coherent. Across all pairs, sentences contained on average  $9.3 \pm 1.7$  words (range 2–24), and a story contained on average  $14.9 \pm 6.8$  sentences.

### Stability and Entrainment

**Figures 1C,D** show the ITI and IWI data from an individual story. Both the IWIs and ITIs varied a lot from one word to the next, often by several seconds, making word timing unstable. However, the inter-turn intervals of the two participants (1D) were highly correlated (for this example  $r = 0.72$ ,  $p < 0.001$ ) with each other, indicating high entrainment between the participants.

The circular histograms in **Figure 2** confirm this pattern for the whole experiment, demonstrating that word rhythms were highly entrained even though the individual and joint timings were unstable. The distribution of the relative phase angles (**Figure 2**) calculated from the ITIs has a clear preferred direction toward  $180^\circ$ , indicating anti-phase entrainment. The entrainment measure for the pooled data was  $\bar{R} = 0.70$ , and  $\bar{R} = 0.74 \pm 0.05$  for the 16 individual stories. Instead of varying evenly within its observed range ( $14\text{--}326^\circ$ ), the distribution shows a heavy weighting to anti-phase angles, with 95% of the values concentrated between  $78$  and  $270^\circ$ . This phase attraction toward

the anti-phase was also demonstrated in statistical tests, where, in all stories, the observed distributions deviated statistically significantly from uniform distributions ( $p < 0.01$ ; Kuiper test).

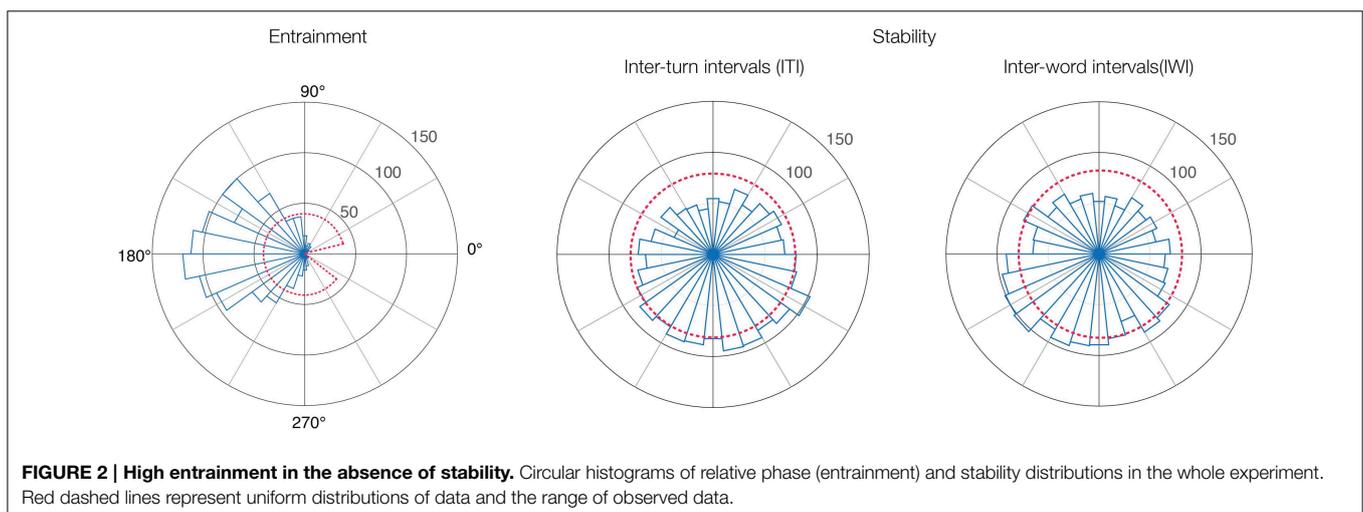
In contrast, the distributions for both the individual word timings (ITI, **Figure 2**) and in the joint time series of word onsets (IWI) were uniformly spread around the circle without any clear preference. The stability measures were very low,  $\bar{R} = 0.14$  for ITI, and  $\bar{R} = 0.13$  for IWI.

Looking at the 16 joint time series and the 32 individual time-series (16 stories \* 2 participants) separately, the stability measure  $\bar{R}$  was  $0.15 \pm 0.064$  for the joint timings (IWIs) and  $0.18 \pm 0.071$  for the individual timings (ITIs). *V*-tests confirmed that with the exception of three cases, all individual ITI time-series were unstable, as the distributions did not differ from uniform distribution toward the expected mean direction of zero at  $p < 0.05$ .

The stability and entrainment scores did not differ between the audio-only and audiovisual conditions ( $p = 0.39$  for IWI,  $p = 0.15$  for ITI, and  $p = 0.15$  for entrainment; paired two-tailed *t*-tests).

## Discussion

We found that when two participants were creating stories together, in turns, one word at a time, their word rhythms were strongly entrained. Such a high level of entrainment was unexpected, as the word rhythms themselves were very unstable, and the participants were not given any instructions related to word rhythm, tempo, or timings of their words. Previously, entrainment of comparable strength has been observed in finger-tapping tasks, where the entraining beats occur at equal intervals and the participants are specifically asked to aim for accurate anti-phase timing. The unexpected independence of high levels of entrainment from a stable word-to-word rhythm is in line with the oscillation-based theory of turn-taking (Wilson and Wilson, 2005), which assumes that conversation participants are entrained to a common rhythm that is established by shared syllable timing (Street, 1984). This shared rhythm governs the



participants' "readiness" to take turns, and it helps them to optimize turn-taking so that it does not comprise overlaps and long silences.

Interpersonal coordination and adaptation occur in a wide range of tasks, such as pronouncing letters of the alphabet (Kawasaki et al., 2013) or in anti-phase finger tapping (Nowicki et al., 2013). These rather simple tasks mainly recruit automatic entrainment processes, whereas our task of joint story building required advanced cognitive operations to guarantee that the story evolved in a meaningful and smooth manner.

As an automatic and subconscious process, entrainment is assumed to subserve communicative interaction (Gallese, 2001, 2005; Himberg, 2014), and in our task, participants needed to be aligned at the semantic and syntactic levels, as well as the speech-process level (Clark, 1996; Garrod and Pickering, 2004). The high entrainment that we observed could be what allowed the participants to reach this multi-level, multimodal coordination (Dale et al., 2013).

In our study, stability and entrainment were statistically similar in "telephone-like" trials (with only auditory connection between the participants) and "video-call-like" trials (with auditory and visual connection between the participants). This result partly agrees with the results of a previous corpus study of face-to-face as well as telephone dialogs, where pause durations between participants were highly correlated in both types of conversations, suggesting entrainment to a common rhythm even in telephone-mediated conversations (Ten Bosch et al., 2004). However, in the corpus study, the pauses were longer and more variable in the face-to-face conversations.

We did not observe such differences, possibly because due to the simultaneous MEG recording, our participants were asked to sit still, which limited the amount and utility of gestural communication between them. All task-critical information was delivered through the auditory channel.

Although our participants conducted the story-building task in a laboratory setting that restricted their body movements, highly entrained speech rhythms emerged spontaneously in their interaction. This mutual adaptation of speech rhythms implies speech as a strong inducer of entrainment, even when the participants just hear each other.

## Acknowledgments

This study was supported by the European Research Council (Advanced Grant #232946 to RH), Academy of Finland grants #131483 and #263800, Finnish Graduate School of Neuroscience, and the Doctoral Programme Brain and Mind. We thank Pamela Baess, Mathieu Bourguignon, Mia Illman, Veikko Jousmäki, Jyrki Mäkelä, Jussi Nurminen, Lauri Parkkonen, Petteri Räisänen, Ronny Schreiber, and Andrey Zhdanov for contributing to the realization of the 2-person MEG experiments, during which these behavioral data were collected.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00797/abstract>

## References

- Abney, D. H., Paxton, A., Dale, R., and Kello, C. T. (2014). Complexity matching in dyadic conversation. *J. Exp. Psychol. Gen.* 143, 2304–2318. doi: 10.1037/xge0000021
- Baess, P., Zhdanov, A., Mandel, A., Parkkonen, L., Hirvenkari, L., Mäkelä, J. P., et al. (2012). MEG dual scanning: a procedure to study real-time auditory interaction between two persons. *Front. Hum. Neurosci.* 6, 83. doi: 10.3389/fnhum.2012.00083
- Bavelas, J., Black, A., Lemery, C., and Mullett, J. (1986). I show how you feel: motor mimicry as a communicative act. *J. Pers. Soc. Psychol.* 50, 322–329. doi: 10.1037/0022-3514.50.2.322
- Bernieri, F. J., and Rosenthal, R. (1991). "Interpersonal coordination: behavior matching and interactional synchrony," in *Fundamentals of Nonverbal Behavior*, eds R. S. Feldman and B. Rimé (Cambridge: Cambridge University Press), 401–432.
- Boker, S., Covey, E., Tiberio, S. and Deboeck, P. (2005). "Synchronization in dancing is not winner-takes-all: Ambiguity persists in spatiotemporal symmetry between dancers," in *Proceedings of the North American Association for Computational, Social, and Organizational Science* (Notre Dame, IN).
- Chartrand, T., and Bargh, J. (1999). The chameleon effect: the perception-behavior link and social interactions. *J. Pers. Soc. Psychol.* 76, 893–910. doi: 10.1037/0022-3514.76.6.893
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Clayton, M., Sager, R., and Will, U. (2004). In time with the music: The concept of entrainment and its significance for ethnomusicology. *ESEM Counterpoint 1*, 1–45. Available online at: <http://oro.open.ac.uk/id/eprint/2661>
- Condon, W., and Ogston, W. (1967). A segmentation of behavior. *J. Psychiatr. Res.* 5, 221–235. doi: 10.1016/0022-3956(67)90004-0
- Cowley, S. J. (1998). Of timing, turn-taking, and conversations. *J. Psycholinguist. Res.* 27, 541–571. doi: 10.1023/A:1024948912805
- Cummins, F. (2009). Rhythm as entrainment: the case of synchronous speech. *J. Phon.* 37, 16–28. doi: 10.1016/j.wocn.2008.08.003
- Dale, R., Fusaroli, R., Duran, N., and Richardson, D. C. (2013). The self-organization of human interaction. *Psychol. Learn. Motiv.* 59, 43–95. doi: 10.1016/b978-0-12-407187-2.00002-2
- Dixon, N. F., and Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception* 9, 719–721. doi: 10.1068/p090719
- Fisher, N. (1993). *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.
- Gallese, V. (2001). The shared manifold hypothesis. From mirror neurons to empathy. *J. Conscious. Stud.* 8, 33–50.
- Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenol. Cogn. Sci.* 4, 23–48. doi: 10.1007/s11097-005-4737-z
- Garrod, S., and Pickering, M. (2004). Why is conversation so easy? *Trends Cogn. Sci.* 8, 8–11. doi: 10.1016/j.tics.2003.10.016
- Giles, H., Coupland, N., and Coupland, J. (1991). "Accommodation theory: communication, context, and consequence," in *Contexts of Accommodation: Developments in Applied Sociolinguistics*, eds H. Giles, N. Coupland, and J. Coupland (Cambridge: Cambridge University Press), 1–68.
- Gonzales, A. L., Hancock, J. T., and Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communic. Res.* 37, 3–19. doi: 10.1177/0093650209351468
- Himberg, T. (2006). "Co-operative tapping and collective time-keeping - differences of timing accuracy in duet performance with human or computer partner," in *Proceedings of the ICMPC 9*, eds M. Baroni, A. R. Addressi, R. Caterina, and M. Costa (Bologna), 377.
- Himberg, T. (2014). *Interaction in Musical Time*. Doctoral dissertation, Faculty of Music, University of Cambridge, Cambridge.

- Himberg, T., and Thompson, M. R. (2011). Learning and synchronising dance movements in South African songs – cross-cultural motion-capture study. *Dance Res.* 29, 305–328. doi: 10.3366/drs.2011.0022
- Hove, M. J., and Risen, J. L. (2009). It's all in the timing: interpersonal synchrony increases affiliation. *Soc. Cogn.* 27, 949–960. doi: 10.1521/soco.2009.27.6.949
- Iivonen, A. (1998). "Intonation in Finnish," in *Intonation Systems: A Survey of Twenty Languages*, eds D. Hirst and A. Di Cristo (Cambridge: Cambridge University Press), 311–327.
- Jammalamadaka, S. R., and Sengupta, A. (2001). *Topics in Circular Statistics*, Vol. 5. Singapore: World Scientific Publishing.
- Jansen, J., and Bulterman, D. C. (2013). "User-centric video delay measurements," in *Proceedings of the 23rd ACM Workshop on Network and Operating Systems Support for Digital Audio and Video* (Oslo: ACM), 37–42.
- Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., and Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Sci. Rep.* 3:1692. doi: 10.1038/srep01692
- Kendon, A. (1970). Movement coordination in social interaction: some examples described. *Acta Psychol.* 32, 101–125. doi: 10.1016/0001-6918(70)90094-6
- Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol.* 63, 2220–2230. doi: 10.1080/17470218.2010.497843
- Kurita, T., Lai, S., and Kitawaki, N. (1994). Effects of transmission delay in audiovisual communication. *Electron. Commun. Jpn.* 77, 63–74. doi: 10.1002/ecja.4410770306
- Louwerse, M. M., Dale, R., Bard, E. G., and Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cogn. Sci.* 36, 1404–1426. doi: 10.1111/j.1551-6709.2012.01269.x
- Lucas, G., Clayton, M., and Leante, L. (2011). Inter-group entrainment in Afro-Brazilian congado ritual. *Empir. Musicol. Rev.* 6, 75–102. Available online at: <http://hdl.handle.net/1811/51203>
- Malloch, S., and Trevarthen, C. (2009). "Musicality: communicating the vitality and interests of life," in *Communicative Musicality - Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 1–11.
- Nessler, J. A., and Gilliland, S. J. (2009). Interpersonal synchronization during side by side treadmill walking is influenced by leg length differential and altered sensory feedback. *Hum. Mov. Sci.* 28, 772–785. doi: 10.1016/j.humov.2009.04.007
- Nowicki, L., Prinz, W., Grosjean, M., Repp, B. H., and Keller, P. E. (2013). Mutual adaptive timing in interpersonal action coordination. *Psychomusicol. Music Mind Brain* 23, 6–20. doi: 10.1037/a0032039
- Pikovsky, A., Rosenblum, M., and Kurths, J. (2001). *Synchronization - a Universal Concept in Non-linear Sciences* (No. 12). Cambridge: Cambridge University Press.
- Rankin, S. K., Large, E. W., and Fink, P. W. (2009). Fractal tempo fluctuation and pulse prediction. *Music Percept.* 26, 401–413. doi: 10.1525/mp.2009.26.5.401
- Reich, C. M., Berman, J. S., Dale, R., and Levitt, H. M. (2014). Vocal synchrony in psychotherapy. *J. Soc. Clin. Psychol.* 33, 481–494. doi: 10.1521/jscp.2014.33.5.481
- Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992. doi: 10.3758/BF03206433
- Richardson, M. J., Marsh, K. L., and Schmidt, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 62–79. doi: 10.1037/0096-1523.31.1.62
- Richardson, M., Marsh, K., Isenhower, R., Goodman, J., and Schmidt, R. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Mov. Sci.* 26, 867–891. doi: 10.1016/j.humov.2007.07.002
- Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 696–735. doi: 10.1353/lan.1974.0010
- Scott, S. K. (1998). The point of p-centres. *Psychol. Res.* 61, 4–11. doi: 10.1007/PL00008162
- Shockley, K., Santana, M.-V., and Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 326–332. doi: 10.1037/0096-1523.29.2.326
- Sidnell, J. (2001). Conversational turn-taking in a Caribbean English Creole. *J. Pragmatics* 33, 1263–1290.
- Stivers, T., Enfield, N., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106
- Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Hum. Commun. Res.* 11, 139–169. doi: 10.1111/j.1468-2958.1984.tb00043.x
- Ten Bosch, L., Oostdijk, N., and De Ruiter, J. P. (2004). "Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues," in *Proceedings of the 7th International Conference Text, Speech and Dialogue*, eds P. Sojka, I. Kopeček, and K. Pala Brno (Berlin: Springer).
- Toivola, M., Lennes, M., and Aho, E. (2009). "Speech rate and pauses in non-native Finnish," in *Proceedings of the 10th Annual Conference of the International Speech Communication Association*. (Brighton).
- Vos, P. G., Mates, J., and van Kruysbergen, N. W. (1995). The perceptual centre of a stimulus as the cue for synchronization to a metronome: evidence from asynchronies. *Q. J. Exp. Psychol.* 48, 1024–1040. doi: 10.1080/14640749508401427
- Vroomen, J., and Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: not that special. *Cognition* 118, 75–83. doi: 10.1016/j.cognition.2010.10.002
- Wilson, M., and Wilson, T. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 12, 957–968. doi: 10.3758/BF03206432
- Wiltermuth, S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5. doi: 10.1111/j.1467-9280.2008.02253.x
- Zhdanov, A., Nurminen, J., Baess, P., Hirvenkari, L., Jousmäki, V., Mäkelä, J. P., et al. (in press). An internet-based real-time audiovisual link for dual meg recordings. *PLoS ONE*.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Himberg, Hirvenkari, Mandel and Hari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.