# Burstiness across multimodal human interaction reveals differences between verbal and non-verbal communication

**Drew H. Abney (dhabney@indiana.edu)**
Department of Psychological and Brain Sciences
Indiana University, Bloomington, IN 47405 USA

**Rick Dale (rdale@ucmerced.edu)**
**Chris T. Kello (ckello@ucmerced.edu)**
Cognitive and Information Sciences
University of California, Merced, Merced, CA 95340 USA

**Max M. Louwerse (m.m.louwerse@uvt.nl)**
Center for Cognition and Communication
Tilburg University, Tilburg, Netherlands

**Abstract**

Recent studies of naturalistic face-to-face communication have demonstrated temporal coordination patterns such as the synchronization of verbal and non-verbal behavior, which provides evidence for the proposal that verbal and non-verbal communicative control derives from one system. In this study, we argue that the observed relationship between verbal and non-verbal behaviors depends on the level of analysis. In a re-analysis of a corpus of naturalistic multimodal communication (Louwerse et al., 2012), we focus on measuring the temporal patterns of specific communicative behaviors in terms of their *burstiness*. We examined burstiness estimates across different roles of the speaker and different communicative channels. We observed more burstiness for verbal versus non-verbal channels, and for more versus less informative language sub-channels. These findings demonstrate a new method for analyzing temporal patterns in communicative behaviors, and they suggest a more complex relationship between verbal and non-verbal channels than suggested by prior studies.

**Keywords:** burstiness, multimodal communication, verbal and non-verbal communication

## Introduction

In cognitive science, a considerable number of studies have investigated the role of non-verbal communication in relation to verbal communication. The majority of these studies suggest an intrinsic relationship between verbal and non-verbal communication. For instance, a strong link has been shown between lexical access and gesturing, such that when people gesture, lexical access is facilitated (Rime & Schiaratura, 1991). Also, the time gap between gesture and a familiar word is considerably shorter than the gap between gesture and an unfamiliar word (Morrel-Samuels & Krauss, 1992), and when speech is disrupted, gestures are halted (Mayberry & Jaques, 2000). Gesture is thought to be intrinsically related to language processing (Butterworth & Morrissette, 1996) because most gestures occur when people speak (McNeill, 1992), and because of evidence linking gesture with language development (Butcher & Goldin-Meadow, 2000). In fact, non-verbal and verbal communication are sometimes argued to be so interwoven that gesture and speech are co-expressive manifestations of one integrated system, forming complementary components of one underlying process that helps organize thought (Goldin-Meadow, 2005; McNeill, 1992).

Louwerse, Dale, Bard, and Jeuniaux (2012) investigated the temporal relationship between matching behaviors in dialog partners, such as manual gesture in one speaker vs. the same manual gesture in the other speaker. By applying a cross-recurrence analysis, Louwerse et al. showed synchronized matching behavior in all categories (language, facial, gestural) that were investigated at temporal lags short enough to suggest imitation of one speaker by the other. Louwerse et al. concluded that the similarities between the different channels – verbal and non-verbal – demonstrated that the temporal structure of matching behaviors provided low-level and low-cost resources for human interaction.

So far, all studies focusing on the similarities between verbal and non-verbal communication, including Louwerse et al. (2012), focused on the *temporal matching* of verbal and non-verbal behavior. They tend not to investigate the *temporal distribution* of independent behavioral event dynamics. Complex behaviors such as human interaction tend not to show the strictest forms of synchrony, but instead are more loosely, functionally coupled (e.g., Fusaroli et al., 2014). Instead, the overall pattern of behavior, expressed in the distribution of events, may reflect particular local patterns of interaction – when one interlocutor gestures, it may sustain itself for a given period of time before waning; when another person speaks, this burst of behavior may look quite different, sustaining itself for longer, more regular periods of time. These event dynamics might paint a different picture of the relationship between verbal and non-verbal channels.

## The Property of Burstiness

Most work studying human communication is based on dyadic analyses that focus on temporal patterns across partners rather than the temporal patterns of specific behaviors produced by each partner. In the current study, the large multimodal corpus of human communication collected and reported in Louwerse et al. was re-analyzed to focus on the quantification of a particular property of behavior, burstiness.

Using the framework developed by Goh and Barabasi (2008) and extended by others (e.g., Jo, Karsai, Kertész, & Kaski, 2012), we estimated the burstiness of verbal and non-verbal behaviors. The burstiness parameter, $B$, provides an estimate of a system's activity patterns spanning from periodic ($B = -1$), to random ($B = 0$), to theoretically maximal burstiness ($B = 1$) (see Figure 1). Goh and Barabasi (2008) observed that human phenomena like human texts and email patterns have positive burstiness estimates, $B > 0$, whereas human cardiac rhythms were found to have periodic burstiness estimates, $B < 0$.

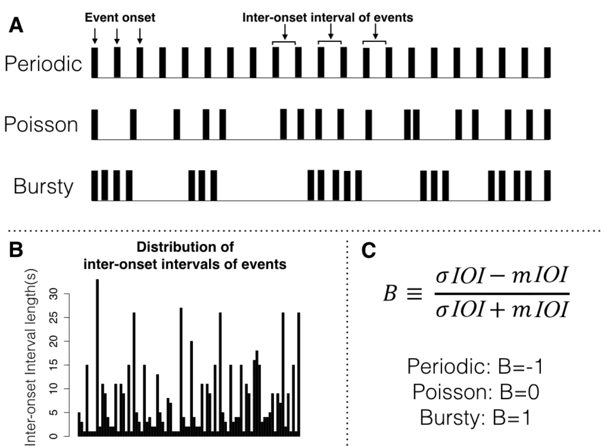



Figure 1: Overview of system's activity patterns spanning from periodic, to random, to theoretically optimally bursty.

## The Current Study

The goal of the current study was to investigate the temporal dynamics of behavioral events across verbal and non-verbal communicative modalities during face-to-face human interaction. We focus on the measure of burstiness, now widely used in statistical physics to capture the temporal patterns of point processes in complex network interactions.

In the first analysis section, we investigated whether or not there were differences in the burstiness of behaviors that are categorized into verbal and non-verbal channels. It is possible that verbal and non-verbal channels have similar degrees of burstiness, which would be consistent with previous work suggesting a strong intrinsic relationship. However, if the channels exhibit different degrees of burstiness, such results would suggest a more complex relationship between verbal and non-verbal communication. To further explore and understand the burstiness measure, we also investigated the burstiness of sub-channels that constitute the language communicative channel. Our results indicate that burstiness is different for verbal versus non-verbal behaviors, and also for different aspects of language behaviors.

# Methods

## Multimodal Communication Corpus

The original task developed to collect these multimodal data is described by Louwerse et al. (2008) and Louwerse et al. (2012), who were interested in collecting multimodal structure of human interaction in order to inform avatar design for intelligent tutoring systems and other technologies. In the task, $N = 24$ pairs of participants helped each other navigate a map. Each pair of participants completed 8 rounds of navigation. For each round, one participant was chosen as the "Information Giver", and other the "Information Follower." The Information Giver had a complete map, and the Information Follower had a noisy and partial map. This mismatch between maps was intended to elicit communication and predict the points at which misunderstandings were likely to occur. The participants had to use language and gesture via webcam so that the Information Follower could reconstruct a map route with the help of the Information Giver. The corpus was developed by taking these 192 recordings of interactions and coding a wide variety of behaviors. These codings were based on well-known or adapted coding schemes in discourse, along with some other semi-automated procedures (see Louwerse et al., 2008 for details). All behaviors were coded in 250ms to encompass relatively fast behaviors such as nodding, acknowledgements, and smiling. The output from this coding procedure was a multicolumnar data format of binary point series that represented the occurrence of different behaviors at a 250ms interval. These 250ms intervals were the subject of our burstiness analyses.

We chose 39 behaviors that fit into four specific behaviors channels (as did Louwerse et al., 2012). Behavioral channels were categorized into two factors, Channel and Role. For the Channel factor, channels were identified as either "Face & Head," "Manual Gesture," "Face Touch," or "Language." For the Role factor, channels were identified as either Giver or Follower. For the levels of the Role factor, all channels were included for the Giver and the Follower. See Table 1 for the behaviors that were included into each channel. The language sub-channels were annotated at the utterance-level.

Table 1: List of Channels, Sub-channels, and Behaviors

| channel | sub-channels | behaviors |
|---|---|---|
| face & head | mouth | laughing, lip tightening |
| | eyes | blink, rolling eyes |
| | eyebrows | asymmetrical, down-frowning, out brow raiser |
| | head | nodding, shaking |
| manual gesture | | beat, deictic, iconic, metaphoric, symbolic |
| touch face | | touching cheek, chinrest |
| language | dialogue acts | acknowledgements, align, check, clarify, explain, instruct, query-what, query-yes/no, ready, reply-no, reply-what, reply-yes |
| | discourse connectives | alright, no, ok, um, well, yes |
| | descriptions | color, compass direction, digit, relative direction, spatial preposition |

## Construction of Multivariate Spike Trains and Inter-event Intervals

We are interested in estimating the burstiness of multimodal communicative behavior and are therefore working with a multivariate class of *spike trains*. To our knowledge, the current study provides the first steps towards dealing with burstiness in multivariate spike train corpora. The protocol converts multivariate spike trains into inter-event interval (IEI) distributions. These interval distributions help quantify the temporal clustering of communicative events across channels.

First, for each behavior, we created a spike train of onset events which excludes successive '1's for prolonged events. Second, for each communicative channel (Face & Head, Manual Gesture, Face Touch, Language), we summed the spike trains from each behavior, yielding a multimodal event series where a '0' represents a sample when no event occurred, a '1' represents a sample when one event occurred, and any number greater than 1 represents a sample when two or more events occurred. For example, a sample with a "Laughing" event and a "Nodding" event would have a "2" in the event series. Any sample with two or more events is considered a sample of simultaneous communicative behavior which we discuss below. Finally, IEI's were computed from the multimodal event series to construct an IEI distribution for each channel for each map task role (Giver or Follower).

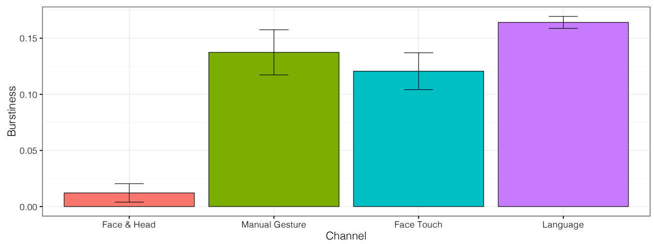
An IEI is computed by considering two consecutive events, $t_j$ and $t_{j+1}$, and finding the temporal difference between them, $\tau = t_{j+1} - t_j$. For an IEI that contains simultaneous communicative behavior (2 or more events in the same sample), an IEI, $\tau$, was computed and added to the distribution in addition to a zero for each additional event. For example, when an IEI with the second sample has 3 events, we would add to the IEI distribution (1) the corresponding $\tau$ and (2) two zeros (0,0). We chose to add this component to the protocol because we wanted to treat simultaneous communicative behavior as quantitatively 'more bursty'. Adding zeros to an IEI distribution will amplify a burstiness estimate. IEI distributions for each communicative channel and each map task role were submitted to estimates of burstiness.

## Estimation of Burstiness

The burstiness parameter, $B$, is defined as,

$$B = \frac{\sigma_\tau - m_\tau}{\sigma_\tau + m_\tau}$$

where $\sigma_\tau$ is the standard deviation of the IEI distribution and $m_\tau$ is the mean of the IEI distribution (Goh & Barabási, 2008; Jo, Karsai, Kertész, & Kaski, 2012). Alternative measures of burstiness have been employed in previous studies in computational linguistics (Altmann, Pierrehumbert, & Motter, 2009; Pierrehumbert, 2012) utilizing parameter fitting from a stretched exponential distribution (Weibull distribution). These alternative measures have provided unique insights into the dynamics of linguistic levels of description. Our decision to utilize the burstiness parameter, $B$, is twofold. First, parameter estimation from a distribution requires a minimum number of data points or IEIs. Therefore, with the properties of our corpus, parameter estimation from distribution fitting requires the implementation of confidence intervals, which can be avoided with the utilization of the burstiness parameter, $B$. Second, one goal of this study is to account for simultaneous communicative behavior as a higher degree of burstiness. The burstiness parameter, $B$, is amplified when zeros are added to the IEI distribution and therefore an ideal option for the current study. $B$ is bounded from [-1,1], where $B = 1$ for a theoretical maximum bursty behavior, $B = -1$ for completely regular behavior (e.g., metronome), and $B = 0$ for a homogeneous Poisson process, i.e., independent events. We omitted trials that did not include reliable burstiness estimates for any of the four channels across the MapTask roles in the first analysis section (1.24% of trials) and for any of the three channels across the MapTask roles in the second analysis section (1.00% of trials).



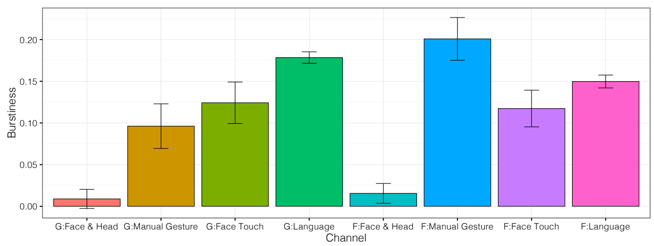




Figure 2a and 2b: Burstiness across channels with a) Information Giver (G) and Follower (F) combined, and b) the roles separated. Error bars reflect 95% CIs.

## Investigating Differences in Burstiness across Verbal and Non-verbal Channels

Mixed effects models (Bates et al., 2014; Team R., 2013) were utilized to determine if burstiness differed across different channels. The first set of analyses was conducted to compare burstiness estimates across role structure and communicative channels. Linear models were utilized to predict burstiness estimates. Fixed effects for these models included map task role (leader or follower), communicative channels (Face & Head, Manual Gesture, Face Touch, and Language), and event count for each communicative channel. Event count was added into the model as a covariate to control for the potential relationship between burstiness estimates and the number of behavioral events going into the analysis. Dyad and map type were included as random effects.

If there are differences across communicative channels, we can observe such differences in a variety of ways: are there differences in the temporal structure across communicative modalities (1) collapsing burstiness estimates across MapTask roles? (2) within MapTask roles (e.g., Follower:Manual Gesture vs. Follower:Language)? and/or (3) across MapTask roles (e.g., Follower:Manual Gesture vs. Giver: Manual Gesture)?

Collapsing burstiness estimates across MapTask role, we observed a significant main effect of communicative channel, $F(3, 1030) = 162.55$, $p < .0001$ (Figure 2a). See Table 2 for results from multiple comparison tests. Overall, the language channel ($M$=.16, $SE$=.003) was observed to be more bursty relative to the manual gesture channel ($M$ = .14, $SE$ = .01), $b$ = .08, $p$ = .009.

Table 2: Multiple Comparisons from the random mixed effects model: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

| | Multiple Comparisons | Beta | Z-score |
|---|---|---|---|
| Channel | Man. Gest. v. Face/Head | .08 | $7.9^{***}$ |
| | Touch Face v. Face/Head | .11 | $9.6^{***}$ |
| | Language v. Face/Head | .17 | $16.7^{***}$ |
| | Touch Face v. Man. Gest. | .02 | 2.2 |
| | Language v. Man. Gest. | .08 | $7.5^{***}$ |
| | Language v. Touch Face | .05 | $4.6^{***}$ |
| Role | Leader v. Follower | .01 | .7 |
| Int. | F:Man. Gest v. F:Lang | -.07 | $-4.79^{***}$ |
| | G:Man. Gest. v. G:Lang | .05 | $3.17^{*}$ |
| | F:Man. Gest v. G:Lang | .06 | $3.08^{*}$ |
| | G:Man. Gest v. F:Lang | .04 | 2.79 |

The communicative channel x map task role interaction was significant, $F(3, 1030) = 20.97$, $p < .0001$, therefore, we tested for multiple comparisons using Tukey Honestly Significant Difference tests to investigate differences within and across MapTask roles (Figure 2b). At this level of the analysis, we were specifically interested in the differences between language and manual gestures, so we limit our report to those subsets of the analysis. We observed within-role differences between language and manual gesture burstiness estimates for the Follower role ($b$ = -.07, $p$ < .001) and for the Giver role ($b$ = .05, $p$ = .03). We also observed a between-role difference for Follower: Manual Gesture v. Giver: Language ($b$=.06, $p$=.04). The results from this analysis suggest that, across map task role, the verbal channel (i.e., language channel) had higher burstiness estimates relative to the non-verbal channels, and specifically the manual gesture channel.

## Investigating the Relative Magnitude of Burstiness in the Language Channel

In the last section, we established that communicative channels exhibit temporal patterns of behavior that (1) vary across verbal and non-verbal channels and (2) are all bursty relative to exhibiting random or periodic temporal patterns. But what does it mean to be *more* bursty? It is important to note that these channels are made up from specific sub-channels that are further made up from individual behaviors. In an effort to better understand the relative magnitude of burstiness, in this section, we focused on the language channel because this channel exhibited the highest estimates of burstiness. Specifically, we zoomed into the language channel and investigated the temporal patterns of the sub-channels.

The language channel is made up of three specific sub-channels: dialogue acts, discourse connectives, and descriptions. We expected to observe higher burstiness estimates for the 'descriptions' sub-channel relative to the other two channels. This hypothesis is motivated by previous research that focused on the burstiness of various linguistic levels in texts (Altmann, Cristadoro, & Esposti, 2012; Altmann, Pierrehumbert, & Motter, 2009). Altmann et al. (2009) observed that burstiness increased across semantic classes where 'entities' like proper nouns had higher burstiness estimates relative to predicates like *in*, which in turn had higher estimates than higher level operators like *the*. If the results observed in texts are consistent with human dialogue, we should expect to observe that descriptions like providing a relative direction will have higher burstiness estimates relative to dialogue acts like saying *no* or discourse connectives like saying *um*.

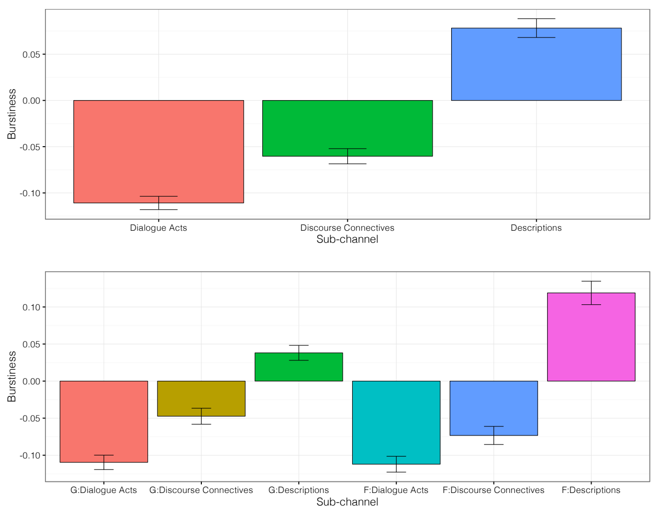



Figure 3a and 3b:
Burstiness across language channels with a) Information Giver (G) and Follower (F) combined, and b) the roles separated. Error bars reflect 95% CIs.

Linear models were utilized to predict burstiness estimates. Fixed effects for these models included map task role (Giver or Follower), language sub-channels (Dialogue Acts, Discourse Connectives, Descriptions), and event count for each communicative channel. Similar to the previous analysis section, event count was added into the model to act as a covariate to control for the potential relationship between burstiness estimates and the number of events going into the analysis. Dyad and map type were again included as random effects. We observed that descriptions ($M$ = .08, $SE$ = .005) had higher burstiness estimates relative to discourse connectives ($M$ = -.06, $SE$ = .004, $b$ = .06, $p <$ .001) and dialogue acts ($M$ = -.11, $SE$ = .004; $b$ = .17, $p <$ .001) (Figure 3a). Discourse connectives and dialogue acts were both more periodic than bursty, and dialogue acts were more periodic (closer to -1) relative to discourse connectives ($b$ = .11, $p <$ .001). These results suggest that various levels of verbal dialogue have different temporal patterns and such patterns have interesting parallels to previous research studying the burstiness of text corpora. We discuss these parallels in addition to the insights gained from the analysis section to better understand the pattern of results in the previous analysis section.

# Discussion

The primary goal of the current paper was to better understand the temporal patterns of verbal and non-verbal behaviors during face-to-face multimodal human communication. We submitted the multimodal corpus to an analysis of burstiness. In the first analysis section, we observed that communicative channels differed in the degree of burstiness, with the verbal channel having higher burstiness estimates relative to non-verbal channels like manual gestures, face & head, and face touch. To add nuance to this result, in the second analysis section, we focused on better understanding the magnitude of burstiness, and zoomed into the language channel. In this analysis, we observed that a more informative sub-channel, 'descriptions', had higher burstiness estimates relative to sub-channels that focused on operators and modifiers.

Much work in the cognitive sciences has argued that verbal and non-verbal behaviors are intrinsically related via the same communicative system (Golden-Meadow, 2005; McNeill, 1992). Recent work (Louwerse et al., 2012) has made this argument by focusing on evidence of synchronization across verbal and non-verbal channels. In the current paper, we observed that, verbal and non-verbal channels differ in terms of estimates of their temporal burstiness. An important question is what these differences reflect. To begin to find an answer to this question, we examined certain language sub-channels and found higher degrees of burstiness for descriptive productions compared to pragmatic productions like dialog acts or connectives.

Considering the latter results, there are a few possible explanations for the observation that verbal and non-verbal channels exhibit different types of temporal patterns, with the verbal channel exhibiting higher burstiness estimates. The first possible explanation is that increased estimates of burstiness for the verbal channel means that more information is contained within this communicative channel relative to the non-verbal channels. This suggestion is influenced by the observations of higher degrees of burstiness in higher-level semantic classes in texts (Altmann, et al., 2009) and higher degrees of burstiness in descriptive sub-channel in dialogue (the current paper's second analysis section). If this is the case, our results point to the proposal that verbal channels during human communication are more informative relative to non-verbal channels. However, this possibility seems unlikely because our own results show that the direction of burstiness estimate differences for the language and manual gesture channels are not consistent: higher estimates for language relative to manual gesture for the information giver and higher estimates for manual gesture relative to language for the information follower.

The second possible explanation is that an important property of multimodal communication is having a collection of different types of temporal patterns across communicative channels. This proposal, what we call the 'temporal heterogeneity' hypothesis, suggests that successful communication emerges from a diverse suite of information channels that vary in temporal properties. An important adaptive property of a complex system, such as a dyadic communicative system (Dale, Fusaroli, Duran, & Richardson, 2013; Fusaroli, Raczaszek-Leonardi, & Tylén, 2013), is the ability for multiple components with specific intrinsic properties to self-organize to form higher-level structures (Kello & Van Orden, 2009; Kugler & Turvey, 1987). This proposal is amenable to the hypothesis that verbal and non-verbal channels are part of the same integrated system (Golden-Meadow, 2003; McNeill, 1992) and that gesture and speech are complementary communicative channels important for the resolution of referential expressions (Louwerse & Bangerter, 2010; Seyfeddinipur & Kita, 2001). The current paper contributes to this line of argument by showing, at a specific level of analysis, that verbal and non-verbal channels have different types of temporal patterns and that the heterogeneity of these temporal patterns might be important for successful communication. Another important contribution is the introduction to a simple analysis of the temporal structure of behavioral event dynamics, the burstiness analysis. Future work is required to better understand the connection between varying degrees of burstiness across diverse types of human behavioral patterns.

## Acknowledgments

This research was supported by NIH, T32 Grant # HD07475 awarded to the first author, and NSF-IIS-0416128 awarded to the last author. The usual exculpations apply.

## References

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLoS One*, *4*(11), e7678.

Altmann, E. G., Cristadoro, G., & Degli Esposti, M. (2012). On the origin of long-range correlations in texts. *Proceedings of the National Academy of Sciences*, *109*(29), 11582-11587.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7).

Butcher, C., & Goldin-Meadow, S. (2000). Language and gesture. *Gesture and the transition from one-to two-word speech: When hand and mouth come together*, 235-257.

Butterworth, G., & Morissette, P. (1996). Onset of pointing and the acquisition of language in infancy. *Journal of Reproductive and Infant Psychology*, *14*(3), 219-231.

Dale, R., Fusaroli, R., Duran, N., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of learning and motivation*, *59*, 43-95.

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

Goh, K. I., & Barabási, A. L. (2008). Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, *81*(4), 48002.

Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think*. Harvard University Press.

Jo, H. H., Karsai, M., Kertész, J., & Kaski, K. (2012). Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics*, *14*(1), 013055.

Kello, C. T., & Van Orden, G. C. (2009). Soft-assembly of sensorimotor function. *Nonlinear dynamics, psychology, and life sciences*, *13*(1), 57.

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Routledge.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive science*, *36*(8), 1404-1426.

Louwerse, M. M., & Bangerter, A. (2010). Effects of ambiguous gestures and language on the time course of reference resolution. *Cognitive Science*, *34*(8), 1517-1529.

Mayberry, R., & Jaques, J. (2000). Gesture production during stuttered speech: insights into the nature of speech-gesture integration. *Language and Gesture*, 199-215.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 615.

Pierrehumbert, J. B. (2012). Burstiness of verbs and derived nouns. In *Shall We Play the Festschrift Game?* (pp. 99-115). Springer Berlin Heidelberg.

Rime, B., & Schiaratura, L. (1991). Gesture and Speech. Chicago.

Seyfeddinipur, M., & Kita, S. (2001, June). Gestures and self-monitoring in speech production. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 27, No. 1, pp. 457-464).