COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2018) 1–20
Copyright © 2018 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12612

The Bursts and Lulls of Multimodal Interaction: Temporal Distributions of Behavior Reveal Differences Between Verbal and Non-Verbal Communication

Drew H. Abney,[a] Rick Dale,[b] Max M. Louwerse,[c] Christopher T. Kello[d]

[a]*Department of Psychological and Brain Sciences, Indiana University*
[b]*Department of Communication, University of California, Los Angeles*
[c]*Cognitive Science and Artificial Intelligence, Tilburg University*
[d]*Cognitive and Information Sciences, University of California, Merced*

Received 7 July 2017; received in revised form 22 November 2017; accepted 26 January 2018

Abstract

Recent studies of naturalistic face-to-face communication have demonstrated coordination patterns such as the temporal matching 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 reanalysis of a corpus of naturalistic multimodal communication (Louwerse, Dale, Bard, & Jeuniaux, 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 modalities. We observed more burstiness for verbal versus non-verbal channels, and for more versus less informative language subchannels. Using this new method for analyzing temporal patterns in communicative behaviors, we show that there is a complex relationship between verbal and non-verbal channels. We propose a "temporal heterogeneity" hypothesis to explain how the language system adapts to the demands of dialog.

Keywords: Multimodal interaction; Verbal communication; Nonverbal communication; Temporal distributions; Burstiness

1. Introduction

Human communication includes a rich, dynamic organization across multiple modalities. Verbal behavior is produced and perceived in fractions of a second, along with non-verbal behaviors that can coincide with it. In cognitive science, a considerable number of

Correspondence should be sent to Drew H. Abney, Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th St., Bloomington, IN 47405. E-mail: dhabney@indiana.edu

studies have investigated the role of non-verbal communication in relation to verbal communication. The majority of these studies suggest an intrinsic relationship between the two. For instance, a strong link has been shown between lexical access and gesturing, such that when people gesture, lexical access is facilitated (Rimé & 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). When speech is disrupted, gestures are halted (Mayberry & Jaques, 2000). Gesture is thought to be intrinsically related to language processing (Butterworth & Morissette, 1996) because most gestures occur when people speak (McNeill, 1992), and because of evidence linking gesture with language development (Butcher & Goldin-Meadow, 2000). Gesture not only operates as a context to language, or language as a context to gesture, but they complement one another (Louwerse & Bangerter, 2010). In fact, non-verbal and verbal behavior 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; Iverson & Thelen, 1999; McNeill, 1992).

Louwerse et al. (2012) investigated the temporal relationship between matching behaviors in dialog partners, such as manual gesture in one speaker versus the same manual gesture in the other speaker, and showed behavior matching (language, facial, gestural) at temporal lags short enough to suggest synchronization of one speaker by the other. Louwerse et al. (2012) concluded that the number of modalities showing this behavior-matching synchronization, the short temporal lags, and 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 communication, serving "as an active and adaptive background process supporting an interactive task" (p. 18).

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. Temporal matching can be defined as when a specific behavior is produced by two interlocutors within a specific temporal window. They tend not to investigate the *temporal distribution* of independent behavioral event dynamics. Complex behaviors such as human interaction typically do not show strict synchrony. Instead, they are more loosely, functionally coupled (e.g., Abney, Paxton, Dale, & Kello, 2015; Fusaroli, Rączaszek-Leonardi, & Tylén, 2014; Wallot, Mitkidis, McGraw, & Roepstorff, 2016). Moreover, the overall pattern of behavior, expressed in the distribution of events, may reflect particular local patterns of interaction. For example, as one interlocutor *gestures*, that gesturing may sustain itself for a given period of time, with periodic onsets of activity before waning. When another interlocutor *speaks*, this burst of behavior may look quite different, sustaining itself for longer, but with clustered onsets of behavior. This pattern of timing is not simply synchrony, but rather encompasses the relative stochastic distribution of verbal and non-verbal events during an interaction. By focusing on the temporal distributions rather than matching, these event dynamics might paint a different picture of the relationship between verbal and non-verbal channels than what the cognitive science literature has reported so far. In the current study, we investigated the temporal distributions of verbal and non-verbal communicative channels during a dyadic interaction. Specifically, we submitted a multimodal corpus of dyadic interaction to an analysis of burstiness, providing a simple measure of the temporal distributions of specific behavioral events.

1.1. Burstiness

We focus on burstiness, a property widely used in statistical physics to capture the temporal patterns of point processes in complex network interactions (Goh & Barabási, 2008; Holme & Saramäki, 2012; Karsai et al., 2011), but relatively new to the cognitive sciences. Most work in cognitive science 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. Of course, it is insightful to know how coordination, alignment, and synchronization emerge across speakers, but such findings would benefit from knowing how modalities distribute temporally within a speaker. For instance, knowing how modalities temporally match can inform us how speakers coordinate their conversation (Louwerse et al., 2012). Indeed, emerging findings continue to provide new insights into how humans temporally coordinate with each other across a variety of contexts (Abney et al., 2015; Fusaroli & Tylén, 2016; Fusaroli et al., 2012, 2014; Louwerse et al., 2012; Paxton & Dale, 2013a,b; Schmidt, Nie, Franco, & Richardson, 2014; Shockley, Baker, Richardson, & Fowler, 2007; Shockley et al., 2007; Tolston, Shockley, Riley, & Richardson, 2014; Wallot et al., 2016). However, little is known about the temporal distributions of individual modalities within a speaker during human interaction. Uncovering the temporal distributions of specific behaviors produced by individuals during an interaction will provide a better understanding of the temporal ensemble of behavioral signals that encompass multimodal communication. In the current study, the large multimodal corpus of human communication collected and reported in Louwerse et al. (2012) was reanalyzed to focus on the quantification of a particular property of behavior produced by each partner, that is, burstiness.

Using the framework developed by Goh and Barabási (2008) and extended by others (e.g., Jo, Karsai, Kertész, & Kaski, 2012), we estimated the burstiness of verbal and nonverbal 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 Fig. 1). Goh and Barabási (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, $B < 0$. In this framework, a system displaying *periodic,* homogenous timing activity patterns produces equally spaced behavioral events. A system displaying *bursty*, non-homogeneous timing activity patterns produces, behavioral events that are temporally clustered followed by longs lulls of inactivity. Human phenomena have been observed to produce non-homogeneous timing patterns (e.g., Barabási, 2005; Malmgren, Stouffer, Motter, & Amaral, 2008; Vásquez et al., 2006). Many have speculated that the underlying mechanism that generates bursty, non-homogeneous behaviors is either a combination of behaviors occurring at different periodicities across multiple timescales, the interdependencies of

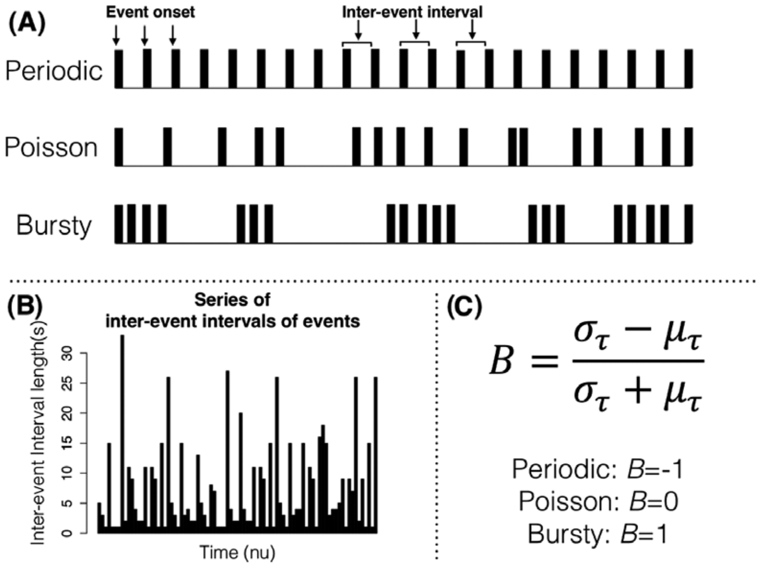

Fig. 1. Overview of computing burstiness estimates from interevent interval distributions derived from event series. (A) Examples of three event series exhibiting (top) periodic, (middle) Poisson, and (bottom) bursty temporal structure. (B) An interonset interval (IEI) time series. (C) The equation for calculating burstiness estimates from IEI distributions.

behavioral events, or a combination of processes (Anteneodo, Malmgren, & Chialvo, 2010; Jung, Jang, Kralik, & Jeong, 2014; Malmgren et al., 2008). Despite the open question about the underlying mechanisms of non-homogeneous timing patterns, these processes are considered to be a property of a complex, dynamical system (Barabási, 2005; Karsai, Kaski, Barabási, & Kertész, 2012).

The burstiness framework is particularly useful for the current study's purposes because it (a) provides an index of the temporal structures of verbal and non-verbal communicative behavior estimated from a distribution of interevent intervals of behavioral events and (b) can also provide a simplified view into the non-homogeneous processes that are features of complex, dynamical systems. Although there have been many documented observations of non-homogeneous processes generating bursty or clustered dynamics in human communicative behavior (e.g., Abney, Paxton, Dale, & Kello, 2014; Altmann, Cristadoro, & Degli Esposti, 2012; Falk & Kello, 2017; Kello et al., 2017), there is also evidence that the temporal structure of communicative behavior is rhythmic and periodic (Cummins & Port, 1998; Dauer, 1983; Kohler, 2009; Tilsen & Arvaniti, 2013). It is, therefore, an open question regarding where verbal and non-verbal communicative behaviors fall on a spectrum of temporal structure.

1.2. The current study

The goal of the current study was to investigate the temporal distributions of behavioral events across verbal and non-verbal communicative modalities *within* a speaker during face-to-face collaborative human interaction. In the first analysis section, considering the theoretical (Goldin-Meadow, 2005; Iverson & Thelen, 1999; McNeill, 1992) and empirical (Louwerse et al., 2012) arguments that suggest verbal and non-verbal behaviors are generated from one integrated system, we investigated whether or not there were differences in the burstiness of behaviors that are categorized into verbal and non-verbal modalities. It is possible that verbal and non-verbal modalities have similar temporal distributions, such as either both modalities exhibiting bursty temporal patterns or both modalities exhibiting periodic (i.e., rhythmic) patterns, in addition to exhibiting indistinguishable degrees of distributional patterns. For example, it is possible that verbal and non-verbal behaviors both show periodic temporal patterns and these patterns, as measured by the burstiness metric, are equivalent to one another. We call this hypothesis, the "temporal homogeneity" hypothesis. It is also possible that the modalities have the same types of temporal distributions but have different degrees of distributional patterns, for example, *more bursty* or *more periodic*. This alternative hypothesis, what we call the "temporal heterogeneity" hypothesis, would suggest a more complex relationship between verbal and non-verbal channels. The temporal heterogeneity hypothesis is consistent with a view of complex communicative systems where multiple components have distinct intrinsic properties—such as different temporal processes—that self-organize to adapt to changing external constraints and contexts (Dale, Fusaroli, Duran, & Richardson, 2013; Fusaroli, Rączaszek-Leonardi, & Tylén, 2013; Kello & Van Orden, 2009; Kugler & Turvey, 1987).

In the first analysis section we investigated whether or not there were differences in the burstiness of verbal and non-verbal behavior in general. In the second analysis section, we investigated the burstiness of subchannels that constitute the language communicative modality.

2. Methods

The methods of this study are described in Louwerse et al. (2012) and are summarized below.

2.1. Participants

Forty-eight students (24 dyads; 30 females and 18 males; 19 African American, 1 Asian, and 28 Caucasian) from the University of Memphis participated in this study for payment. All participants were native speakers of English.

2.2. Multimodal communication corpus

The original task developed to collect these multimodal data was described by Louwerse, Jeuniaux, Zhang, Wu, and Hoque (2008) and Louwerse et al. (2012), who collected multimodal structure of human interaction in order to inform avatar design for intelligent tutoring systems and other technologies. Each participant sat at a computer and communicated with his or her partner via video-conferencing software. Therefore, the experimental setup afforded a virtual face-to-face interaction while completing a spatial orienting task. Cameras and microphones (bird's eye view, face, torso) recorded the participants' verbal and non-verbal behavior. Each pair of participants completed eight rounds of navigation. For each round, one participant was chosen as the "Instruction Giver" (IG) and other the "Instruction Follower" (IF), with each participant being IG (and IF) four rounds. The IG had a complete map, and the IF had a noisy and partial map. The task for the IG was to use her complete map to navigate the IF through her noise and partial map. The mismatch between maps was intended to elicit communication and predicted the points at which misunderstandings were likely to occur. An example of such a map task and sample maps is provided in Fig. 2.

The corpus was developed by taking these 192 recordings of interactions and coding a wide variety of behaviors ensuring intercoder reliability. These codings were based on well-known or adapted coding schemes in discourse, along with some semi-automated procedures (see Louwerse et al., 2008, 2012, for details). All coded actions were polled at 250-ms intervals, as shorter time intervals seemed to be undesirable with many of the actions occurring at a second interval (e.g., gestures, dialog acts), and longer time intervals seemed to be undesirable with a relatively fast occurrence (e.g., nodding, acknowledgments, smile).

The output from this coding procedure was a multicolumnar data format of binary point series that represented the occurrence of different behaviors at a 250-ms interval. From the range of modalities described in Louwerse et al. (2012) we selected language and manual gesture (Table 1). These 250-ms intervals of the 29 behaviors in these two modalities were the subject of our burstiness analyses.

Two factors that were used to divide behaviors: Role and Modality. For the Role factor, manual gesture and language modalities were identified as either Instruction Giver (IG) or Instruction Follower (IF). As discussed in the next subsection, specific behaviors were superimposed into a behavioral event series specific to Role and Modality. Therefore, there were four specific behavioral event series, (a) IG: Manual Gesture, (b) IG: Language, (c) IF: Manual Gesture, and (d) IF: Language.

3. Analyses

3.1. Construction of multivariate spike trains and interevent intervals

Our aim was to estimate the burstiness of multimodal communicative behavior and we are, therefore, working with a multivariate class of *spike trains*. The current study provides the first steps toward analyzing burstiness in multivariate multimodal 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 within and across modalities.

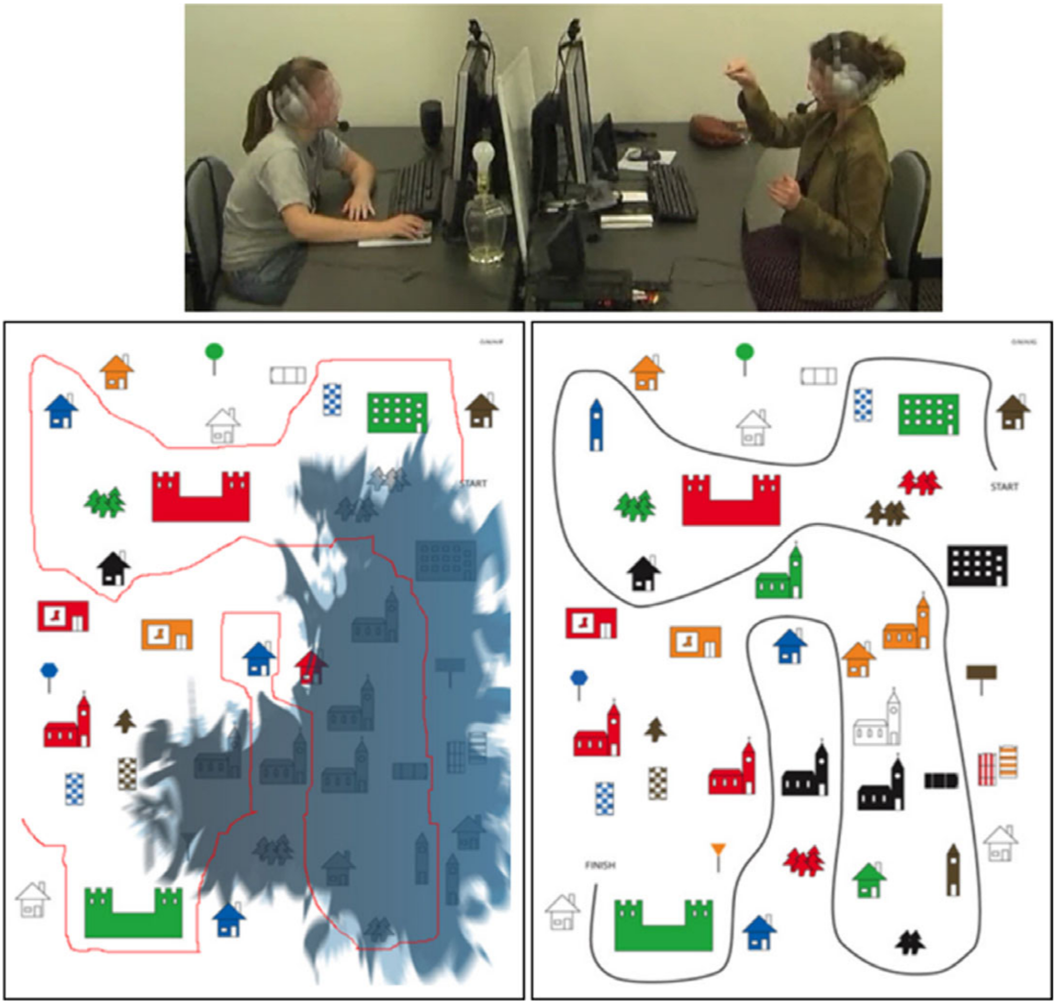

Fig. 2. Sample maps used in the experiment. The map for the Instruction Giver is on the right, and the map for the Instruction Follower (with the route drawn by her) is on the left. The inkblot on the Instruction Follower's map is an experimental manipulation that prevents the Instruction Follower to see color and thus to elicit dialog.

First, for each behavior, we created a spike train of onset events, which excluded successive "1"s for prolonged events. For instance, an iconic gesture that lasted 750 ms was included only by its onset. Second, for each communicative modality (Manual Gesture, Language), we superimposed 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, for the Manual Gesture modality, we superimposed five specific spike trains that corresponded to the behaviors identified as

Table 1
List of modalities, subchannels, and behaviors

Modality	Subchannels	Behaviors
Manual gesture		Beat, deictic, iconic, metaphoric, symbolic
Language	Dialog acts	Acknowledgments, 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

manual gesture behaviors (see Table 1). The superposition of behaviors into a multivariate spike train made it possible to determine if behaviors occurred simultaneously. For example, a 250-ms sample with the discourse connective "um" used by one interlocutor, coinciding with the discourse connective "well" in that same 250-ms sample for the same interlocutor, would generate the value "2" in the event series. Any sample with two or more events is considered a sample of intrapersonal simultaneous communicative behavior to be discussed below. Finally, IEIs were computed from the multimodal event series to construct an IEI distribution for each modality (Manual Gesture or Language) for each role (IG or IF).

An IEI was 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 contained simultaneous communicative behavior (two or more events in the same sample), an IEI of $\tau = 0$ was added for each simultaneous event beyond the first. For example, when there were three events at time t_j, such that t_j, t_{j+1}, and t_{j+2} are all equal, two IEIs of zero were added to the IEI distribution. This method of analysis treats simultaneous communicative behavior as quantitatively "more bursty" because adding zeros to an IEI distribution, by nature of the burstiness equation, amplifies its burstiness estimate. Moreover, the addition of zeros made the simplifying assumption that all events within a temporal window occurred simultaneously. A more detailed approach might assume that events are distributed evenly within a temporal window, which would instead create $n-1$ IEIs each of duration $dt/(n-1)$, where dt is the minimal temporal window (e.g., 250 ms) and n is the number of simultaneous events. The results are nearly identical when employing the more detailed approach. The alternative to these approaches is to not account for the small IEIs between events that occur in the same temporal window and collapse all events to one event. This alternative approach assumes that a system cannot produce more than one event per minimum temporal window, which runs counter to a perspective of a communicative system with modalities producing behavioral events across multiple timescales. IEI distributions for each communicative modality and each map task role (IG or IF) were submitted to estimates of burstiness.

3.2. Estimation of burstiness

The burstiness estimate indexes a property that represents the combination of bursts and lulls of a particular behavior. The burstiness parameter, B, is defined as,

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

where σ_τ is the standard deviation of the IEI distribution and μ_τ is the mean of the IEI distribution (Goh & Barabási, 2008; Jo et al., 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 two-fold. 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 different types of temporal distributions such as periodic, random, or bursty. Other analyses such as Allan Factor, which provides a metric of the nested hierarchical clustering of events, conflates periodic and random structure (Abney et al., 2014). Moreover, 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, that is, independent events. We omitted trials that did not include burstiness estimates due to a zero count in the IEI distribution for any of the two modalities across the map task roles in the first analysis section (19.20% of trials; for example, if the IG or IF did not gesture during the trial) and for any of the three language subchannels across the map task roles in the second analysis section (1.00% of trials). See Fig. 1 for an overview of the implementation of the burstiness metric.

3.3. Simulations for burstiness classification

To calibrate our interpretations of burstiness estimates, in this section, we analyze simulated IEI series to determine the upper and lower bounds of the “random” category given the numbers and rates of events in our corpus. Given that our multimodal corpus had 620 total event streams across 24 dyads, and the average event count for each series was 147 events, we generated 620 IEI series from the exponential distribution with length = 147. An exponential distribution is a probability distribution that describes events that occur independently and will approximate $B = 0$. For each simulated IEI series we calculated an estimate of burstiness. From the distribution of simulated burstiness estimates, we computed 95% confidence intervals. The mean of the simulated burstiness distribution was $B = -0.003$, the upper 95% CI was $B = 0.005$, and the lower 95% CI was $B = -0.01$. The upper and lower 95% CIs were used as cutoffs in subsequent analyses to classify empirical IEIs into the three temporal structure categories: bursty, periodic, or random.

4. Results

4.1. Investigating differences in burstiness across verbal and non-verbal channels

Table 2 and Fig. 3 show the classifications of temporal distributions. Across IG and IF roles, 80% and 99% of all event series were categorized as bursty for the manual gesture and language modalities, respectively. For the IG, 70% of manual gestures event series were categorized as bursty and 27% were categorized as periodic; 100% of the language event series were categorized as bursty. For the IF, 93% of the manual gesture event series were categorized as bursty and 6% were categorized as periodic; 99% of the language event series were categorized as bursty and 1% was categorized as periodic.

These results point to a few important observations. Across and within IG and IF roles, the language modality was almost always categorized as bursty relative to periodic or random. The majority of manual gesture event series were categorized as bursty as well. However, depending on the Map Task role (IG or IF), there were different distributions of periodic classifications and bursty classifications, as demonstrated in Table 2 and Fig. 3.

Linear mixed-effects (LME) models (Bates, Maechler, Bolker, & Walker, 2014; Team R., 2013) were utilized to determine whether estimates of burstiness differed across different modalities. The first set of analyses was conducted to compare burstiness estimates across role structure and communicative modality. LME models were utilized to predict burstiness estimates. Fixed effects for these models included map task role (IG or IF), communicative modalities (Manual Gesture and Language), and event count for each communicative modality. Event counts were added into the model as a covariate to

Table 2
Percentage of trials classified into the three temporal structure categories for the two modalities

	Periodic (%)	Random (%)	Bursty (%)
Manual gesture	19.07	1.27	79.66
Language	0.52	0	99.48
IG			
Manual gesture	27.27	2.1	70.63
Language	0	0	100
IF			
Manual gesture	6.45	0	93.55
Language	1.04	0	98.96

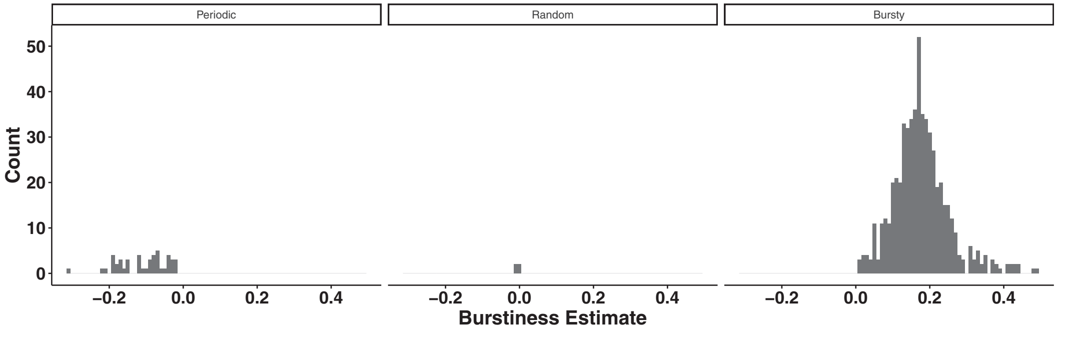

Fig. 3. Histograms of burstiness estimates classified into the three temporal structure categories.

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 interrogate such differences in a variety of ways. For example, are there differences in the temporal structure across communicative modalities collapsing burstiness estimates across map task roles? Are there differences *within* roles (e.g., IF:Manual Gesture vs. IF:Language)? Are there differences *across* roles (e.g., IG:Manual Gesture vs. IF:Manual Gesture)?

Collapsing burstiness estimates across map task role, we observed a significant main effect of communicative modality, $F(1, 424) = 11.72$, $p < .001$, suggesting that the language modality ($M = 0.16$, $SE = 0.003$) was observed to be burstier relative to the manual gesture modality ($M = 0.14$, $SE = 0.01$), $b = 0.06$, $p < .001$, $d = 0.27$. We also observed a significant main effect of Map Task role, $F(1, 424) = 6.36$, $p = .01$, suggesting that the IF ($M = 0.16$, $SE = 0.005$) was observed to be burstier relative to the IG ($M = 0.14$, $SE = 0.006$), $b = 0.11$, $p < .001$, $d = 0.20$. The Communicative Modality $\times$ Map Task Role interaction was significant, $F(1, 424) = 56.61$, $p < .001$, and we, therefore, tested for multiple comparisons using Tukey Honestly Significant Difference tests to investigate differences within and across map task roles and communicative modalities. For the IF, burstiness estimates for manual gestures ($M = 0.20$, $SE = 0.01$) were higher than for the language modality ($M = 0.15$, $SE = 0.004$), $z = -4.22$, $p < .001$, $d = 0.54$. In contrast, for the IG, burstiness estimates for the language modality ($M = 0.18$, $SE = 0.003$) were higher than the estimates for the manual gesture modality ($M = 0.10$, $SE = 0.01$), $z = 3.42$, $p = .003$, $d = 0.67$.

Within-modality differences were found for manual gesture, suggesting that the IF ($M = 0.20$, $SE = 0.01$) had higher burstiness estimates relative to estimates for the IG ($M = 0.10$, $SE = 0.01$), $z = 8.16$, $p < .001$, $d = 0.71$. No within-modality differences were found for language, $p = .34$. Finally, burstiness estimates from the language modality by the IF ($M = 0.15$, $SE = 0.004$) were higher relative to estimates from manual gestures by the IG ($M = 0.10$, $SE = 0.01$), $z = 3.07$, $p < .001$, $d = 0.42$.

The results from this analysis suggest that, in general, the verbal modality had higher burstiness estimates relative to the manual gesture modality (see Table 3 and Fig. 4). The results also suggest that underneath this general pattern is a role-specific effect (Fig. 4), with IF interlocutors having burstier gestures, whereas the main effect for verbal burstiness is driven by the IG. We revisit the implications of these role-specific effects in the Discussion.

Finally, although we did not have any specific predictions of event count, to be conservative with our analyses, we added event count as a covariate in our models to account for any possible influences on the burstiness estimates. Fig. 4 shows how event count differs across modality. Importantly, there are no discernable patterns between event count and the burstiness estimates.

4.2. *Relative magnitude of burstiness in the language modality*

In the previous section, we established that communicative modalities exhibit temporal patterns of behavior that (a) vary across verbal and non-verbal modalities and (b) are overall, burstier relative to exhibiting random or periodic temporal patterns. But what does it mean to be *more* bursty? In an effort to better understand the relative magnitude of burstiness, in this section, we focused on the language modality and its subchannels (and individual behaviors) because this modality exhibited the highest estimates of burstiness.

The language modality as used in this study is made up of three specific subchannels: dialog acts, discourse connectives, and descriptions. We expected to observe higher burstiness estimates for the "descriptions" subchannel 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 et al., 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 dialog, we should expect to observe

Table 3
Multiple comparisons from the random mixed-effects model

	Multiple Comparisons	*b*	*Z*-Score
Within modality			
Language	IF vs. IG	−0.01	−1.62
Manual gesture	IF vs. IG	0.11	8.16***
Across modality			
IF	Manual gesture vs. language	−0.07	−4.22***
IG	Manual gesture vs. language	0.06	3.42**
Interaction			
	IF: manual gesture vs. IG: language	0.05	2.41
	IG manual gesture vs. IF: language	0.04	3.07***

*$p < .05$, **$p < .01$, ***$p < .001$.

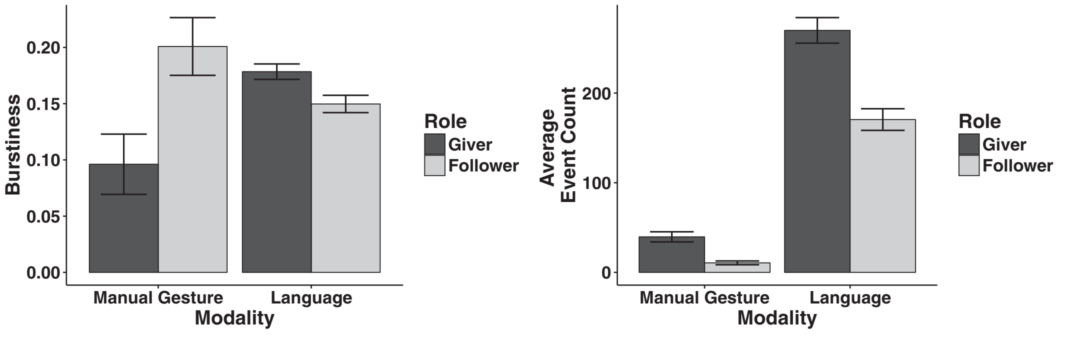

Fig. 4. (Left) Burstiness across modalities and map task role. (Right) Average event count across modality and map task role. Error bars reflect 95% CIs.

that descriptions—like providing a relative direction—will have higher burstiness estimates relative to dialog acts like saying *no* or discourse connectives like saying *um*.

The three subchannels of the language modality showed distinct classification distributions. Over 93% of all "dialogue acts" event series were categorized as periodic. Seventy-three percent of all "discourse connectives" event series were categorized as periodic and 21% were categorized as bursty. A majority (76%) of "descriptions" event series were categorized as bursty. All of these patterns were approximately the same when broken down by map task role (see Table 4).

As stated earlier, LME models were utilized to predict burstiness estimates. Fixed effects for these models included map task role (IG or IF), language subchannels (Dialogue Acts, Discourse Connectives, Descriptions), and event count for each subchannel. Similar to the previous analysis, 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. A significant main effect was found for subchannel, ($F[1, 951] = 621.51$, $p < .001$), suggesting that there were differences in burstiness estimates across the three language subchannels. Descriptions ($M = 0.08$, $SE = 0.005$) had higher burstiness estimates relative to discourse connectives ($M = -0.06$, $SE = 0.004$, $b = 0.06$, $p < .001$, $d = 1.55$) and dialog acts ($M = -0.11$, $SE = 0.004$; $b = 0.17$, $p < .001$, $d = 2.20$). Discourse connectives and dialog acts were both more periodic than bursty, and dialog acts were more periodic (closer to -1) relative to discourse connectives ($b = 0.11$, $p < .001$, $d = 0.67$). These patterns of results are consistent when breaking down the subchannels across map task role (see Fig. 5).

As discussed in the previous section, we did not have any specific hypotheses regarding event count and included the variable into the model as a covariate to control for any effects on the burstiness estimates. Fig. 4 shows for the average event counts for each subchannel and map task role. These results suggest that various levels of verbal dialog have different temporal patterns and such patterns have interesting parallels to previous research studying the burstiness of text corpora.

Table 4
Percentage of trials classified into the three temporal structure categories for the subcategories of the language modality

	Periodic (%)	Random (%)	Bursty (%)
Dialog acts	93.49	1.56	4.95
Discourse connectives	73.70	4.95	21.35
Descriptions	17.59	5.77	76.64
IG			
Dialog acts	93.23	2.60	4.17
Discourse connectives	69.79	4.69	25.52
Descriptions	25.52	6.77	67.71
IF			
Dialog acts	93.75	0.52	5.73
Discourse connectives	77.60	5.21	17.19
Descriptions	9.52	4.76	85.71

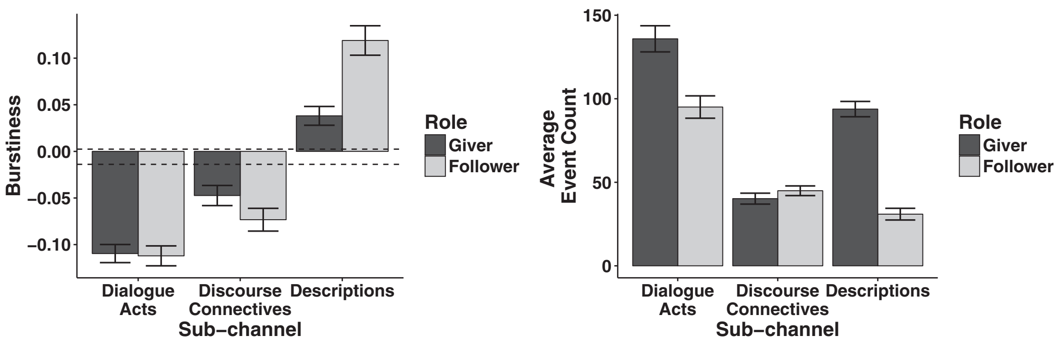

Fig. 5. (Left) Burstiness across subchannels and map task role. Horizontal lines reflect the upper and lower 95% CIs of the "random" category derived from simulations. (Right) Average event count across subchannels and map task role. Error bars reflect 95% CIs.

We now 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.

5. 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 a large corpus of verbal and non-verbal communicative behavior between two participants to an analysis of burstiness. We proposed two hypotheses—the *temporal homogeneity* hypothesis and the *temporal heterogeneity* hypothesis—that present opposing expectations about the relationship between temporal processes across verbal and non-verbal modalities. In the first analysis, we observed that communicative modalities differed in the degree of burstiness, with the verbal modality having higher burstiness estimates relative to the non-verbal modality of manual gestures. Adding some nuance to this result, in the second analysis, we focused on the language modality to better understand what underlies verbal burstiness. Here, we observed that a more informative subchannel, "descriptions," had higher burstiness estimates relative to subchannels that focused on dialog acts and connectives.

Much work in the cognitive sciences has argued that verbal and non-verbal behaviors are intrinsically related via the same communicative system (Goldin-Meadow, 2005; Iverson & Thelen, 1999; McNeill, 1992). Recent work (Louwerse et al., 2012) has made this argument by focusing on evidence of synchronization across verbal and non-verbal modalities. In the current paper, we observed that verbal and non-verbal modalities differ in terms of estimates of their temporal distributions via an analysis of burstiness. An important question is what these differences reflect. Is there a qualitative shift when burstiness estimates go from $B > 0$ to $B < 0$? To begin to find an answer to this question, we examined certain language subchannels and found that descriptive productions were bursty, whereas pragmatic productions like dialog acts and connectives were periodic.

Considering the latter results, there are a few possible explanations for the observation that verbal and non-verbal modalities exhibit different types of temporal patterns, with the verbal modality exhibiting higher burstiness estimates. The first possible explanation is that increased estimates of burstiness for the verbal modality means that more information is contained within this communicative channel relative to the non-verbal modality. Although more work is required to test for possible relationships between burstiness and more concrete measures of information, this interpretation finds some indirect support from 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 subchannel in dialog (the current paper's second analysis section). If this is the case, our results point to the proposal that verbal modalities during human communication are more informative relative to non-verbal modalities. However, this possibility seems unlikely because our own results show that the differences in burstiness for the language and manual gesture modalities are not consistent. Higher estimates were found for language relative to manual gesture for the IG and higher estimates for manual gesture relative to language for the IF. In other words, interactive role modulates temporal event distributions.

The second possible explanation is that an important property of multimodal communication is having a collection of different types of temporal patterns across communicative modalities. This proposal, what we call the "temporal heterogeneity" hypothesis, suggests that successful communication emerges from a diverse suite of information modalities that vary in temporal properties. An important adaptive property of a complex system, such as a dyadic communicative system (Dale et al., 2013; Fusaroli, Rączaszek-Leonardi, & Tylén, 2014), 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 theoretical and empirical arguments that verbal and non-verbal modalities are part of the same integrated system (Goldin-Meadow, 2005; McNeill, 1992; Louwerse et al., 2012) and that gesture and speech are complementary communicative modalities important for the resolution of referential expressions (Louwerse & Bangerter, 2010; Seyfeddinipur & Kita, 2001). Importantly, our results do not directly attend to the question of whether or not verbal and non-verbal behaviors are part of the same integrated system, but rather investigate the temporal distributions of multimodal communication. The current paper contributes to this line of argument by showing, at a specific level of analysis, that verbal and non-verbal modalities have different types of temporal patterns and that the heterogeneity of these temporal patterns might be important for successful communication.

This hypothesis may help explain the way task role modulates temporal event distributions. Observed burstiness is not merely a function of modality. Instead, while for one individual (or map task role) more burstiness may be found in the verbal modality, the interactive goals of a dyad may be more complex; individuals under specific roles can move away from temporal homogeneity in order to support performance. The IF, for example, may use the non-verbal modality to rapidly introduce bursts of information during the exchange, such as when they are seeking information about the spatial map task. The temporal heterogeneity hypothesis encourages future work to unpack how not just an individual but also the dyad itself may organize these distributional patterns flexibly to achieve interactive goals. For example, it would be interesting to index how the temporal distributional properties of two people during an interaction match or mismatch. Previous work has shown that the temporal clustering of vocalizations match above and beyond random pairings during a conversation, a term called complexity matching (Abney et al., 2014). Another avenue for future research would be to determine how indices of temporal distributions of behaviors produced during human interactions relate to the performance on a dyadic task. This question would shed light on the possibility that temporal distributions of behaviors either index, or even perhaps influences, task performance.

We can only speculate about the relationship between burstiness and cognition; however, an intriguing connection can be made between the burstiness of communicative behavior, which is indexing the combinations of rapid bursts of behaviors and long lulls of inactivity, and the collection of experimental studies seeking to better understand the effects of temporal *repetition* and *spacing* of the presentation of words for subsequent word learning and categorization (e.g., Schwab & Lew-Williams, 2016; Vlach, Ankowski, & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008). Recently, Schwab and Lew-Williams (2016) observed in a sample of 2-year-olds that repetition of label–object pairs across successive sentences facilitated the subsequent learning of new referents. Although it is left to be determined whether this combination of short repetition (i.e., bursts) across successive sentences (i.e., lulls) affects attentional, memory, or processing mechanisms, or if this effect generalizes to adult populations; these results do point to a possible function for communicative behaviors that display bursty characteristics.

From long-tailed distributions of recurrence times between letter correspondences from Darwin and Einstein (Oliveira & Barabási, 2005) to email exchanges (Goh & Barabási, 2008), mobile phone communications (Jo et al., 2012), and the lexical stream of online discussion forums (Altmann et al., 2009), our results add to this growing list of communicative phenomena showing similar types of properties suggestive of bursty dynamics. Our addition to this list provides further support for the pervasiveness of non-homogeneous processes in behaviors that span across a variety of communicative media and expands it to include verbal and non-verbal behaviors. Despite this pervasiveness, it will be important for future work to determine whether or not there is variance across the degrees of burstiness estimates that might explain aspects of communication like comprehension, task performance, or affect.

It is important to note that we indexed the temporal patterns of behaviors using the onsets of specific behavioral events. This was motivated by a growing literature utilizing analyses created in physics and network science that require spike trains of a particular phenomenon (Goh & Barabási, 2008; Jo et al., 2012; Karsai et al., 2012). Of course, other properties of behavioral phenomena can likely provide useful information about the differences between verbal and non-verbal communication. For example, it is possible that, in addition to understanding the temporal patterns of event onsets, incorporating analyses that index event durations might provide novel insights.

One of the main goals of the current paper was to classify temporal structure of behavior into periodic and bursty categories, and therefore the burstiness framework was a more suitable candidate relative to other analyses of event structure like the Allan Factor, which has shown to be useful for indexing the multiscale clustering of behavioral events. However, more work is needed to better understand the temporal structure of different levels of verbal and non-verbal behavior. The burstiness framework indexes the temporal distributions of behavior at one timescale: the average timescale that people produce behaviors which approximates the mean event rate. A growing collection of research suggests that the timing of human speech is hierarchically clustered across multiple timescales and that the scaling of the clustering is sensitive to a number of factors (Abney, Warlaumont, Oller, Wallot, & Kello, 2016; Abney et al., 2014; Falk & Kello, 2017; Luque, Luque, & Lacasa, 2015). For example, Abney et al. (2014) observed that the degree of clustering of acoustic onset events across multiple timescales differed across conversational contexts. More recently, Falk and Kello (2017) have shown that the degree of hierarchical structure of acoustic onset events produced by parents when interacting with their infants differed depending on whether they were speaking or singing to their infant. These findings motivate the proposal that a property of human communicative behavior is the relationship between the temporal structure of behaviors across multiple timescales. The current study adds to this line of work by identifying specific levels of language and classifying the temporal structure of each level. One observation from this study is that the event series from subchannels of the language modality displayed different types of temporal distributions. Specifically, the majority of event series from the "dialogue acts" and "discourse connectives" subchannels were classified as periodic, whereas the majority of event series from the "descriptions" subchannel were classified as bursty. This observation is consistent with the "temporal heterogeneity" hypothesis, but adds another layer of speculation about the function of various types of temporal structure across multiple levels of language: Dialog acts and discourse connectives, showing periodic structure, might be the "glue" that binds a dyadic conversation, keeping the tempo and pace of the interaction, while levels of language such as "descriptions" provide bursts of information as needed. Even though future work is required to better understand the connection between varying degrees of burstiness across diverse types of human behavioral patterns, the current study found more burstiness for verbal versus non-verbal channels, and for more versus less informative language subchannels, suggesting a more complex relationship between verbal and non-verbal channels than suggested by prior studies.

Acknowledgments

This research was supported by NIH, T32 Grant # HD07475 awarded to the first author (DHA), and NSF-IIS-0416128 awarded to the third author (MML). We thank Travis J. Wiltshire and two anonymous reviewers for helpful comments on previous versions of this manuscript. The usual exculpations apply.

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, *143*(6), 2304–2315. https://doi.org/10.1037/xge0000021.

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2015). Movement dynamics reflect a functional role for weak coupling and role structure in dyadic problem solving. *Cognitive Processing*, *16*(4), 325–332. https://doi.org/10.1007/s10339-015-0648-2.

Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2016). Multiple coordination patterns in infant and adult vocalizations. *Infancy*, *22*(4), 514–539.

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. https://doi.org/10.1073/pnas.1117723109.

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. https://doi.org/10.1371/journal.pone.0007678.

Anteneodo, C., Malmgren, R. D., & Chialvo, D. R. (2010). Poissonian bursts in e-mail correspondence. *The European Physical Journal B-Condensed Matter and Complex Systems*, *75*(3), 389–394. https://doi.org/10.1140/epjb/e2010-00139-9.

Barabási, A. L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, *435*(7039), 207–211. https://doi.org/10.1038/nature03459.

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. In D. McNeill (Ed.), *Language and gesture* (pp. 235–258). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/cbo9780511620850.015

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. https://doi.org/10.1080/02646839608404519.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171. https://doi.org/10.1006/jpho.1998.0070.

Dale, R., Fusaroli, R., Duran, N., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of Learning and Motivation*, *59*, 43–95. https://doi.org/10.1016/B978-0-12-407187-2.00002-2.

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

Falk, S., & Kello, C. T. (2017). Hierarchical organization in the temporal structure of infant-direct speech and song. *Cognition*, *163*, 80–86. https://doi.org/10.1016/j.cognition.2017.02.017.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931–939. https://doi.org/10.1177/0956797612436816.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157. https://doi.org/10.1016/j.newideapsych.2013.03.005.

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science*, *40*(1), 145–171. https://doi.org/10.1111/cogs.12251.

Goh, K. I., & Barabási, A. L. (2008). Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, *81*(4), 48002. https://doi.org/10.1209/0295-5075/81/48002.

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

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, *519*(3), 97–125. https://doi.org/10.1016/j.physrep.2012.03.001.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11–12), 19–40.

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. https://doi.org/10.1088/1367-2630/14/1/013055.

Jung, K., Jang, H., Kralik, J. D., & Jeong, J. (2014). Bursts and heavy tails in temporal and sequential dynamics of foraging decisions. *PLoS Computational Biology*, *10*(8), e1003759. https://doi.org/10.1371/journal.pcbi.1003759.

Karsai, M., Kaski, K., Barabási, A. L., & Kertész, J. (2012). Universal features of correlated bursty behaviour. *Scientific Reports*, *2*, 397. https://doi.org/10.1038/srep00397.

Karsai, M., Kivelä, M., Pan, R. K., Kaski, K., Kertész, J., Barabási, A. L., & Saramäki, J. (2011). Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, *83*(2), 025102. https://doi.org/10.1103/PhysRevE.83.025102.

Kello, C. T., & Van Orden, G. C. (2009). Soft-assembly of sensorimotor function. *Nonlinear Dynamics, Psychology, and Life Sciences*, *13*(1), 57.

Kello, C. T., Dalla Bella, S., Médé, B. & Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of The Royal Society Interface*, *14*(135), 20170231. https://doi.org/10.1098/rsif.2017.0231

Kohler, K. J. (2009). Rhythm in speech and language. *Phonetica*, *66*(1–2), 29–45. https://doi.org/10.1159/isbn.978-3-8055-9117-1.

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

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. https://doi.org/10.1111/j.1551-6709.2010.01135.x.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x.

Louwerse, M. M., Jeuniaux, P., Zhang, B., Wu, J., & Hoque, M. E. (2008). The interaction between information and intonation structure: Prosodic marking of theme and rheme. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1984–1989). Austin, TX: Cognitive Science Society.

Luque, J., Luque, B., & Lacasa, L. (2015). Scaling and universality in the human voice. *Journal of The Royal Society Interface*, *12*(105), 20141344. https://doi.org/10.1098/rsif.2014.1344.

Malmgren, R. D., Stouffer, D. B., Motter, A. E., & Amaral, L. A. (2008). A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences*, *105*(47), 18153–18158. https://doi.org/10.1073/pnas.0800332105.

Mayberry, R., & Jaques, J. (2000). Gesture production during stuttered speech: Insights into the nature of gesture-speech integration. In D. McNeill (Ed.), *Language and gesture* (pp. 199–215). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511620850.013

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: 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), 615–622. https://doi.org/10.1037/0278-7393.18.3.615.

Oliveira, J. G., & Barabási, A. L. (2005). Human dynamics: Darwin and Einstein correspondence patterns. *Nature*, *437*(7063), 1251–1251. https://doi.org/10.1038/4371251a.

Paxton, A., & Dale, R. (2013a). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods*, *45*(2), 329–343. https://doi.org/10.3758/s13428-012-0249-2.

Paxton, A., & Dale, R. (2013b). Argument disrupts interpersonal synchrony. *The Quarterly Journal of Experimental Psychology*, *66*(11), 2092–2102. https://doi.org/10.1080/17470218.2013.853089.

Pierrehumbert, J. B. (2012). Burstiness of verbs and derived nouns. In D. Santos, K. Linden, & W. Ng'ang'a (Eds.) *Shall we play the Festschrift game?: Essays on the occasion of Lauri Carlson's 60th Birthday* (pp. 99–115). Berlin: Springer. https://doi.org/10.1007/978-3-642-30773-7_8

Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* (pp. 239–284). New York: Press Syndicate of the University of Cambridge.

Schmidt, R. C., Nie, L., Franco, A., & Richardson, M. J. (2014). Bodily synchronization underlying joke telling. *Frontiers in Human Neuroscience*, *8*, 1–13. https://doi.org/10.3389/fnhum.2014.00633.

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, *52*(6), 879. https://doi.org/10.1037/dev0000125.

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

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 201. https://doi.org/10.1037/0096-1523.33.1.201.

Team, R. C. (2013). R: A language and environment for statistical computing.

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, *134*(1), 628–639. https://doi.org/10.1121/1.4807565.

Tolston, M. T., Shockley, K., Riley, M. A., & Richardson, M. J. (2014). Movement constraints on interpersonal coordination and communication. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 1891. https://doi.org/10.1037/a0037473.

Vázquez, A., Oliveira, J. G., Dezsö, Z., Goh, K. I., Kondor, I. & Barabási, A. L. (2006). Modeling bursts and heavy tails in human dynamics. *Physical Review E*, *73*(3), 036127. https://doi.org/10.1103/PhysRevE.73.036127

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 246. https://doi.org/10.1037/a0025260.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013.

Wallot, S., Mitkidis, P., McGraw, J. J., & Roepstorff, A. (2016). Beyond synchrony: Joint action in a complex production task reveals beneficial effects of decreased interpersonal synchrony. *PLoS ONE*, *11*(12), e0168306. https://doi.org/10.1371/journal.pone.0168306.