



## Original Articles

## Hierarchical organization in the temporal structure of infant-direct speech and song

Simone Falk<sup>a,b,c,\*</sup>, Christopher T. Kello<sup>d</sup><sup>a</sup> Ludwig-Maximilians-University, Munich, Germany<sup>b</sup> Laboratoire Parole et Langage, UMR 7309, CNRS / Aix-Marseille University, Aix-en-Provence, France<sup>c</sup> Laboratoire Phonétique et Phonologie, UMR 7018, CNRS / Université Sorbonne Nouvelle Paris-3, Paris, France<sup>d</sup> University of California, Merced, USA

## ARTICLE INFO

## Article history:

Received 3 May 2016

Revised 1 February 2017

Accepted 28 February 2017

Available online 11 March 2017

## Keywords:

Infant-directed communication

Speech and song

Temporal variability

Hierarchical nested clustering

Language development

## ABSTRACT

Caregivers alter the temporal structure of their utterances when talking and singing to infants compared with adult communication. The present study tested whether temporal variability in infant-directed registers serves to emphasize the hierarchical temporal structure of speech. Fifteen German-speaking mothers sang a play song and told a story to their 6-months-old infants, or to an adult. Recordings were analyzed using a recently developed method that determines the degree of nested clustering of temporal events in speech. Events were defined as peaks in the amplitude envelope, and clusters of various sizes related to periods of acoustic speech energy at varying timescales. Infant-directed speech and song clearly showed greater event clustering compared with adult-directed registers, at multiple timescales of hundreds of milliseconds to tens of seconds. We discuss the relation of this newly discovered acoustic property to temporal variability in linguistic units and its potential implications for parent-infant communication and infants learning the hierarchical structures of speech and language.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

Adults provide infants with “exaggerated” sound structure when speaking or singing with them compared to adult communication. For example, pitch is higher, corner vowels are hyperarticulated, pitch range is larger and pauses are longer (e.g., Fernald et al., 1989; Kuhl et al., 1997; Trainor, Clark, Huntley, & Adams, 1997). Preverbal infants are highly attracted to infant-directed (ID) expressions and prefer to listen to infant- over adult-directed (AD) speech and song (Cooper & Aslin, 1990; Pegg, Werker, & McLeod, 1992; Trainor, 1996). There is also evidence that participating in ID conversations boosts infants’ language learning by speeding up vocabulary growth and enhancing speech processing (e.g., Saffran, Aslin, & Newport, 1996; Weisleder & Fernald, 2013). However, some modifications in ID registers and their functions are matters of recent debate. One area of dissent is the function of increased variability in ID registers, as found for example in ID vowel structure (e.g., Eaves, Feldman, Griffiths,

& Shafto, 2016; Martin et al., 2015). Another area is the temporal structure of ID registers (Martin, Igarashi, Jincho, & Mazuka, 2016).

Temporal structure in ID and AD speech is typically investigated by examining durational properties and the timing of speech units at particular timescales such as segments, syllables, phrases or utterances. Studies of ID speech following this approach show mixed evidence. For example, some studies have found longer durations of segments and syllables in ID vs. AD registers (e.g., Albin & Echols, 1996; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013), while other studies failed to do so (Lee, Kitamura, Burnham, & Todd, 2014; Wang, Seidl, & Cristia, 2015). As noted by McMurray et al. (2013), altered temporal characteristics of segments and syllables in ID expressions may be a byproduct of more global temporal prosodic phenomena (e.g., slower tempo, greater phrase-final lengthening, or enhanced stress patterns; Bernstein-Ratner, 1986; Fernald & Simon, 1984; Fernald et al., 1989; Trehub & Trainor, 1998). On the other hand, local phenomena also impact global characteristics. For example, slower cadence in ID expressions has recently been identified as depending on phrase-final lengthening (Martin et al., 2016).

These results reveal the need for examining and understanding temporal variation in ID and AD communication across multiple timescales (e.g., Leong & Goswami, 2015). Moreover, these timescales are nested within each other (e.g. see Cummins, 2002;

\* Corresponding author at: Laboratoire Phonétique et Phonologie, UMR 7018, CNRS / Université Sorbonne Nouvelle Paris-3, 19 Rue des Bernardins, 75005 Paris, France.

E-mail address: [simone.falk@univ-paris3.fr](mailto:simone.falk@univ-paris3.fr) (S. Falk).

Tilsen & Arvaniti, 2013). At shorter timescales, we observe phonemic variations (e.g., ~20–100 ms), which are nested within syllables and words (e.g., ~100–500 ms), which are nested within phrases (e.g., ~500–4000 ms), which are nested within utterances (~1000–6000 ms). As a consequence, rather than examining any one or two particular measures of duration, the present study aims to capture the hierarchical organization of temporal variation in ID versus AD speech across timescales. In particular, we test whether ID temporal variation may serve to emphasize the hierarchical organization of speech.

Our approach is based on recent studies showing that the acoustic energy in speech signals can be expressed as clusters of varying duration that are nested over a range of timescales (Abney, Paxton, Dale, & Kello, 2014; Luque, Luque, & Lacasa, 2015). These clusters emerge when analyzing patterns of “temporal events”, that is, discrete points in time when a significant modulation of acoustic energy occurs. Several temporal events in close vicinity form a cluster which is delineated by gaps in time between more distant events. Several of these clusters will form a new cluster on a larger timescale, and thus, a nested structure of clusters emerges. Abney et al. (2014) quantified the degree of nested clustering in conversational speech using *Allan Factor (AF) analysis* (Allan, 1966; see below). They found robust nested clustering of temporal events in conversation, and the degree of clustering depended on whether the conversation was friendly or contentious (e.g., a debate about abortion). These results indicate that AF analyses of temporal event clustering are sensitive to differences in speech style.

Given these results, we examine in the present study whether ID and AD styles also exhibit differences in their degree of nested clustering of temporal events across timescales. Moreover, if temporal variation in ID expressions has the function to emphasize hierarchical organization, then we expect to find greater nested clustering in ID than in AD expressions. We tested this hypothesis in two major forms of ID communication, ID speech (i.e., story reading) and ID play song. In addition, to elucidate the connection between nested clustering and hierarchical linguistic structure, we measured temporal variability in linguistic units at multiple hierarchical levels, and used regression analysis to relate these measures to nested clustering.

## 2. Methods

### 2.1. Participants

Fifteen native German-speaking mothers (mean age = 31.8 years,  $SD = 3.2$  years) with their infants aged 6 months (9 f, 6 m,  $M = 5.8$  months,  $SD = 0.9$  months) from German households in the Munich area volunteered to participate in the study. Infants were all born on term and showed normal development. Mothers gave informed consent to participate in the study and received a small gift for their participation.

### 2.2. Stimuli and procedure

Mothers read a German variant of the story “Three little pigs” and sang a variant of a popular German play song (“Bibabutze-mann”) in the presence of their infant (ID) or to the experimenter (AD). During ID recordings, infants were seated or lying on their mother’s lap or they were in close vicinity to the mother. During AD recordings, the infant was in another room, either sleeping or being cared for by another person. In ID story reading, a pause was offered to the mothers after half of the stimuli recording to avoid fuzziness of the infant during long stretches of reading. For the analyses in these cases, both parts of the recordings were concatenated. Recordings were done at the mother’s home using an

Audio Technica Lavalier Microphone and a Zoom H4-N recorder at 44,100 Hz and a 24-bit sampling rate.

### 2.3. Analyses

Analyses comprised three main steps: Converting speech recordings to series of temporal events, determining the degree of nested clustering of these events by computing Allan Factor (AF) functions, and comparing AF functions between ID and AD conditions. The process of extracting temporal events is illustrated in Fig. 1. To identify events, we chose peaks in the Hilbert amplitude envelope (Rao, Prasanna, & Yegnanarayana, 2007). Peak events are different from the onset events used by Abney et al. (2014), but both serve to demarcate clustering in acoustic energy—in fact, preliminary analyses showed that the same results are obtained using either type of event.

Each recording (Fig. 1A) was downsampled to 11,025 Hz to remove energy above ~5500 Hz. Waveforms were then passed forwards and backwards through a bank of 4th order Butterworth filters. The lowest filter was <50 Hz, the highest was >4525 Hz, with 14 additional passband filters spanning the intermediate frequencies, each one half octave in width (Drullman, 1995). Filters made envelopes and events interpretable with respect to frequency, and half-octave bands helped even out power across frequencies. The Hilbert envelope was computed for each frequency band (Fig. 1B) and all peaks within  $\pm 10$  ms were extracted (i.e. peak rate was set to 100 Hz). Peak thresholding was done before combining the events of all bandpass signals into one event series (Fig. 1C). The threshold was set for each recording such that ~55 peaks per second were retained, on average. The threshold was chosen to be high enough to yield stable estimates of variances across all the measured timescales (Lowen & Teich, 2005). Moreover, this procedure normalized the number of peaks relative to recording levels. (Results were not sensitive to moderate changes in these parameters, and the same patterns held when the event threshold was set to yield the same number of events for all recordings, i.e. the grand mean for the above analysis.)

Clustering in event series was measured using Allan Factor (AF) variance, which has been similarly used to analyze neuronal spike trains, eye movements, and heart rate (Rhodes, Kello, & Kerster, 2014; Teich & Lowen, 1994; Viswanathan, Peng, Stanley, & Goldberger, 1997). AF variance is computed by tiling a given signal with windows of size  $T$ , and counting the number of events  $N$  within each window  $I$  (see Fig. 1D). AF variance at timescale  $T$  was computed as the average squared difference in counts between adjacent windows, divided by twice the mean count,

$$A(T) = \frac{\langle (N_i - N_{i+1})^2 \rangle}{2\langle N \rangle}.$$

AF variance acts like a coefficient of variance in event counts, but specifically with respect to adjacent time windows. AF variance relates to clustering precisely because of this adjacency—higher variance means counts are not evenly distributed across windows. AF variances were computed over the range of available timescales  $T$ , where  $T$  is varied as a power of 2. The longest timescale that can be measured is limited by time series length, and the shortest by resolution. If there is no clustering of events beyond chance (i.e. events are Poisson distributed), then  $A(T) \sim 1$  for all  $T$ . If events show nested clustering across timescales, then  $A(T)$  should be  $> 1$  and scale up with  $A(T) \sim T^\alpha$ , where  $\alpha > 0$ .

Finally, we measured the degree of nesting using the slope of a regression line fit to the  $A(T)$  function in log-log coordinates. Greater nesting corresponds to steeper slopes. One AF function was computed for each recording, where the largest AF timescale was 1/16th the length of each recording, and each smaller time-

Download English Version:

<https://daneshyari.com/en/article/5041585>

Download Persian Version:

<https://daneshyari.com/article/5041585>

[Daneshyari.com](https://daneshyari.com)