Cognition 163 (2017) 80–86

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

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

Simone Falk [a,b,c,*], Christopher T. Kello [d]

[a] Ludwig-Maximilians-University, Munich, Germany
[b] Laboratoire Parole et Langage, UMR 7309, CNRS / Aix-Marseille University, Aix-en-Provence, France
[c] Laboratoire Phonétique et Phonologie, UMR 7018, CNRS / Université Sorbonne Nouvelle Paris-3, Paris, France
[d] University of California, Merced, USA

ARTICLE INFO

Article history:
Received 3 May 2016
Revised 1 February 2017
Accepted 28 February 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.

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 (S. Falk).

http://dx.doi.org/10.1016/j.cognition.2017.02.017
0010-0277/

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 ("Bibabutzemann") 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 ±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-

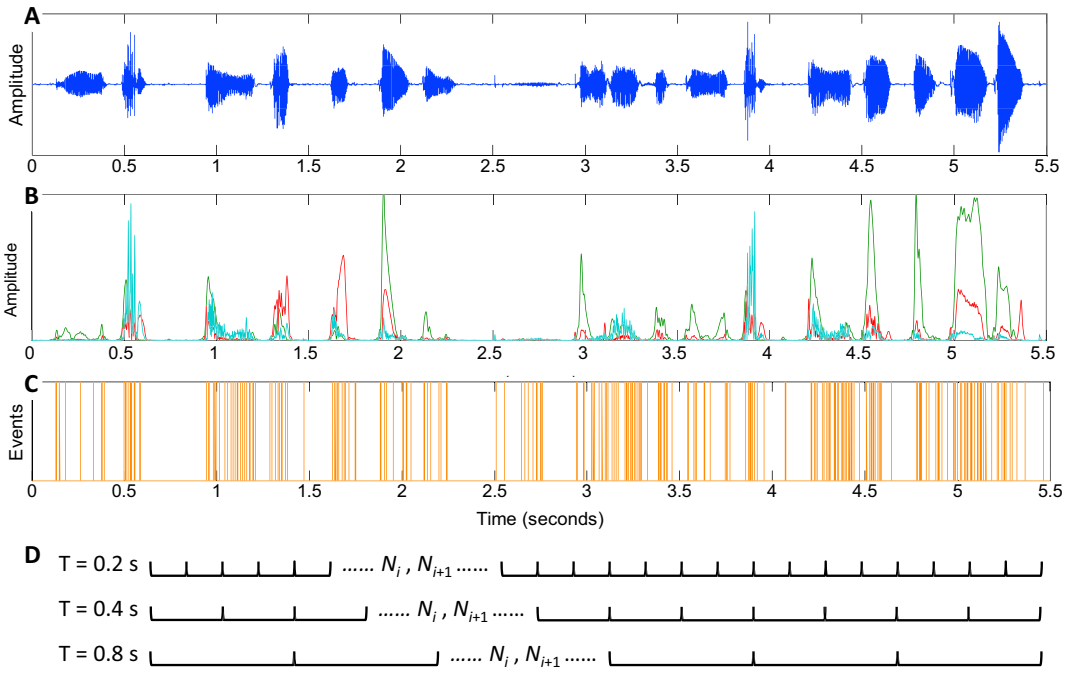

Fig. 1. Example waveform for 5.5 s from one mother reading a story to her infant (A), with corresponding Hilbert envelopes for a sample of three half-octave bands starting at 200 Hz (green), 400 Hz (red), and 800 Hz (turquoise) (B), and peak amplitude events (C). Clustering was measured based on event counts N_i across adjacent windows of size T (three example timescales shown) using Allan Factor variance (see text below) (D).

scale was half the previous one, down to 12 timescales analyzed for each recording.

2.4. Results

ID and AD story recordings were on average 355 and 302 s long, respectively ($t(14) = 4.99$, $p < 0.001$), and ID and AD songs were 270 and 261 s long ($p > 0.05$). Hence, the shortest timescales analyzed was 10 ms and the largest 22 s, on average. An example of both ID and AD event series is given for a story excerpt in Fig. 2.

AF variances and timescales were averaged across participants within each condition, and mean AF functions are plotted in Fig. 3A in log-log coordinates, for story and song recordings in ID and AD conditions. The mean AF functions show little or no clustering below 50 ms ($A(T) \sim 1$), indicating random timing of events at this very small timescale. Clustering beyond chance began around 100 ms as $A(T)$ increased above one and continued to increase across timescales.

Fig. 3A also shows that AF functions differed by AD versus ID register, and speech versus song. Differences began around the 200–500 ms range of timescale, visible in the overall rate of increase in AF variance. Differences were tested by fitting a regression line to each AF function for each recording, and running t-tests on the slopes. Slopes were steeper for ID speech compared with AD speech in story recordings, $M_{ID} = 0.48$ ($CI = \pm 0.069$) vs $M_{AD} = 0.34$ ($CI = \pm 0.035$), $t(14) = 3.95$, $p \sim 0.001$, $d = 1.35$, and in song recordings, $M_{ID} = 0.44$ ($CI = \pm 0.054$) vs $M_{AD} = 0.31$ ($CI = \pm 0.030$), $t(14) = 5.23$, $p < 0.001$, $d = 1.66$. Thus, clustering of events was more nested in ID compared to AD speech and song. Slopes also appeared to be steeper for story-reading compared with singing, but given that the linguistic and phonetic contents were different between these two conditions, we cannot attribute any possible effect to singing or speaking per se.

AF results were consistent with our hypothesis that hierarchical temporal structure is enhanced in ID registers, which can be measured in terms of nested clustering of temporal events in amplitude envelopes. Differences between ID and AD registers spanned a wide range of timescales, suggesting that a number of factors may have contributed to the overall consistent effect on hierarchical temporal structure. Next we investigate some of these factors to gain a better understanding of the main result.

AF variance is affected by the pattern of gaps between event clusters (Abney et al., 2014), which led us to test for potential pause effects that may affect results. First, we ruled out the possibility that the break in ID story-telling caused the results. We repeated AF analysis using only the first halves of each ID and AD recording, and results basically were the same ($M_{ID} = 0.41$ ($CI = \pm 0.043$) vs. $M_{AD} = 0.33$ ($CI = \pm 0.036$), $t(14) = 5.39$, $p < 0.001$, $d = 1.16$). Second, we tested two other possible confounding effects, one deriving from longer pauses in ID vs. AD registers (Fernald et al., 1989), the other from infant sounds during ID speech. We inspected the original sound recordings and removed all pauses longer than 150 ms at intermediate and full intonational phrase boundaries as well as audible infant sounds in the first half of the story reading condition (waveforms were abutted at cut points with a 10 ms pause left in between). This process removed large amounts of recording for three mothers, who were excluded from subsequent analyses (i.e., the remaining recordings were too short). Peak events were extracted and AF analyses done on the remaining trimmed sound data. Results showed the same effects (Fig. 3B): Slopes were steeper for ID versus AD speech, $M_{ID} = 0.33$ ($CI = \pm 0.032$) vs $M_{AD} = 0.26$ ($CI = \pm 0.038$), $t(11) = 3.45$, $p \sim 0.005$, $d = 1.25$.

The observed AF differences support the idea of enhanced hierarchical temporal structure in ID vs. AD registers, but the relation to linguistic hierarchical structure remains unclear. As an initial investigation, we measured temporal variation across several levels of the linguistic hierarchy for ID and AD speech, and investigated whether they explained parts of the variance in ID vs. AD slopes.

As phrase-final lengthening typically differs in ID vs. AD registers (Albin & Echols, 1996; Martin et al., 2016), we chose six disyl-

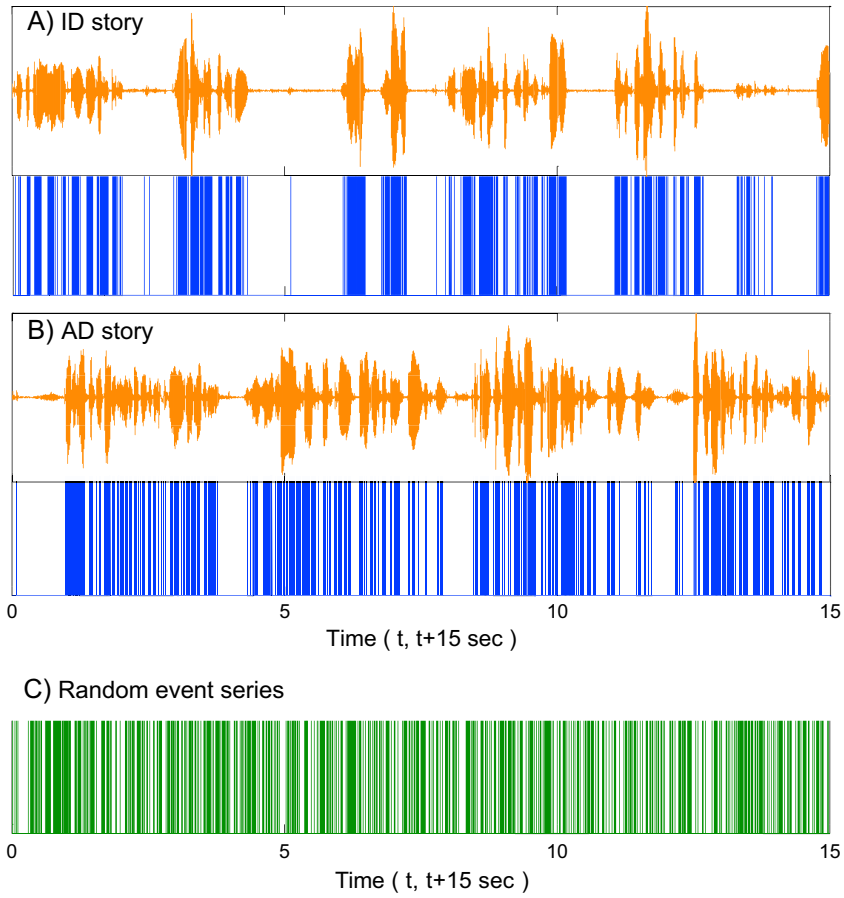

Fig. 2. Example waveforms and event series for one mother telling 15 s (starting at the same point in time) the three little pigs story (German variant). The ID version is displayed on top (A), the AD version underneath (B). Greater clustering can be seen in the ID condition, along with greater nesting i.e. clusters within clusters. At the bottom in green is a randomized (Poisson) event series for comparison, with the same event rate as the data (C).

labic words as test items occurring in utterance-final and non-final positions in each ID and AD story recording (i.e. resulting in 24 items per participant). Associated nested durations (i.e., of the word's stressed vowel ~100 ms, its stressed syllable ~200 ms, the word itself ~400 ms, and of the syntactic phrase in which it occurred ~1500 ms) were segmented using Praat (Boersma, 2001). The coefficient of variation (CV) was computed for each of the four durational measures. In addition, the global mean speech rate of each entire story (syllables/s, including pauses) was calculated as well as a ratio representing pre-final lengthening (PFL, i.e., utterance-final word duration/non-final word duration), and its variability (SD). Overall, there was a tendency for higher temporal variability in ID vs. AD speech, particularly in overall PFL variability and variability at the phrasal constituent level (see Supplementals).

Linear regression models (one model for utterance-final, one for non-final measures) were fitted using difference scores between the ID and AD measures per participant as predictors and difference scores between ID and AD slopes as the dependent variable. Utterance-final predictors from syllable to phrase level (except vowel CV) explained more variance in AF slopes (i.e., 93%, best-fitting model; see Table 1) than non-final predictors. Non-final predictors were not reliably related to AF slopes (in this model, only PFL variability was a significant predictor). These results show that temporal variability related to utterance-final positions at multiple levels of linguistic structure as well as speech rate in ID vs. AD speech jointly contributed to the overall pattern of nested clustering as measured by AF slopes.

3. Discussion

The present findings provide first evidence that hierarchical temporal structure is enhanced in ID registers, and that nested clustering of temporal events corresponds with temporal variability at hierarchical levels of linguistic structure. Using a recently developed method of acoustic analysis, we show that temporal events in the amplitude envelope of ID speech and song exhibit greater clustering than AD expressions, particularly for timescales longer than about 200–500 ms. Initial evidence also indicates that nested clustering in temporal events relates to durational, phrase-final variability in speech at different hierarchical levels, spanning from syllables (about 200 ms) to phrasal constituents (about 1.5 s), and to overall tempo. That is, ID speech and singing to infants at the age of 6 months display more variable durations of utterance-final linguistic units at different time-scales, which, together with global tempo characteristics, appear to generate more nested clustering of acoustic energy in time. Thus, these novel acoustic measures help to effectively reveal clear differences between the rhythmic structures of AD and ID speech and singing which emerge from the interplay of several aspects of temporal variation in speech and song.

Our analysis of variations in linguistic unit durations indicated a primary role for utterance-final variations, which aligns well with previous research on the special role of phrase-final elements in infants' perception of auditory signals. Boundary positions are privileged in attracting infants' attention (Aslin, Woodward, LaMendola, & Bever, 1996; Jusczyk, 1999; Trainor & Adams,

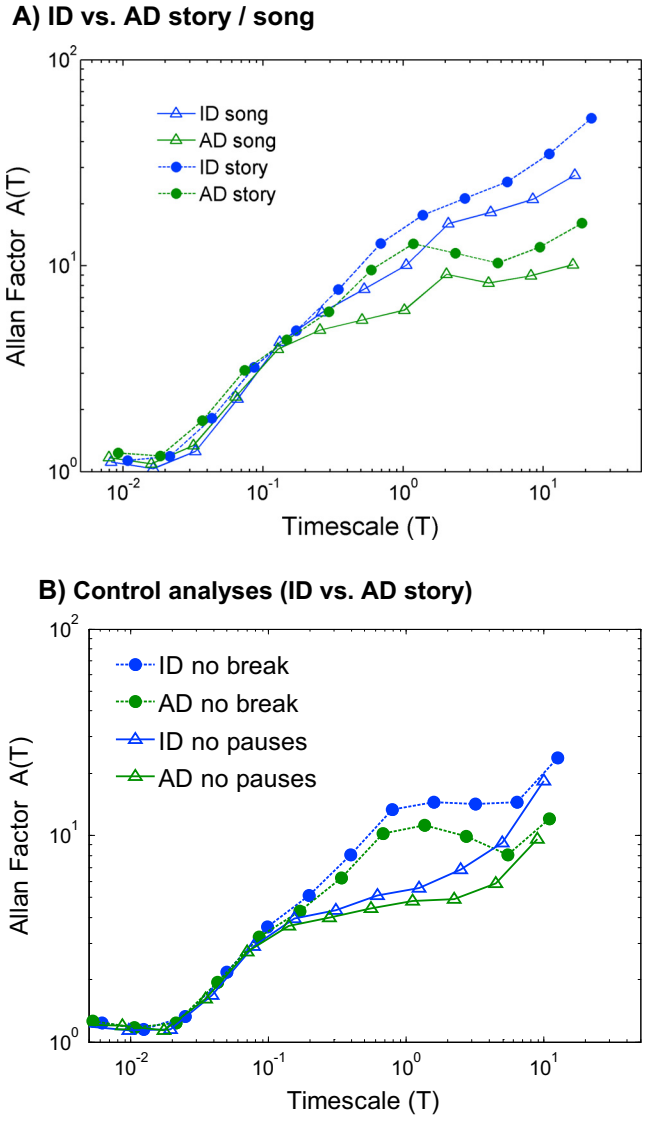

Fig. 3. Panel A: Mean AF functions for ID and AD conditions when mothers were either singing a song or reading a story to their infants, or to an adult (T in seconds). Panel B: Results on ID – AD story reading, when the break between recording sessions was removed from analyses, and when pauses longer than 150 ms and infant sounds were removed. If events occurred randomly (i.e. like a Poisson process), then $A(T)$ functions would be flat at $A(T) = 1$.

2000). Several studies show that infants are more likely to segment and remember words at pre-final boundaries (Seidl & Johnson, 2006; Shukla, White, & Aslin, 2011). As adults tend to place important words and concepts before boundaries, this feature of ID communication is likely to reinforce infant's memory and learning of words (Fernald, 2000; Fernald & Mazzie, 1991). Generally, speakers make prosodic information (such as boundaries and accents) more prominent at times when listeners have minimal chances to foresee the semantic content and information status of upcoming communication (Aylett & Turk, 2004; Kakouros & Räsänen, 2016). Adults may unconsciously follow this principle when communicating with infant listeners who have only marginal knowledge of the ambient language. As an acoustic outcome of these infant and adult preferences, ID communication becomes heavily "bound ary-oriented" (Wang et al., 2015). In spontaneous ID-speech, adults produce *more* prosodic boundaries, especially at higher hierarchical levels (i.e., the utterance; Martin et al., 2016). This leads to the well-described effects that ID utterances are shorter and display more pre-final lengthening, prosodic complexity is reduced (i.e., fewer hierarchically embedded prosodic constituents), and pauses are more frequent than in AD utterances (e.g., Albin & Echols, 1996; Martin et al., 2016; McMurray et al., 2013; Trainor et al., 1997; Wang et al., 2015).

Our study extends these results by showing that greater nested clustering of acoustic energy over time, a novel measure of temporal organization, is also a marker of the boundary-oriented style of ID communication. Temporal clustering is sustained, at least partly, by increased durational variability and contrast of hierarchically nested linguistic units, from the syllable level on. At first glance, this increased variability seems at odds with the idea that ID registers serve the goal of making linguistic content more predictable and accessible to infants. However, Stern (1974) argued that ID communication fulfills the social function to maintain "an optimal range" of infants' attention and arousal during interaction. Infants quickly disengage and lose interest when sensory information becomes too familiar and predictable to them (Sokolov, 1963). Variability and contrast help to make the stimulus more interesting and enhance infants' arousal levels. For example, these effects are clearly visible in the pitch domain, when adults use expanded pitch contours in ID speech to arouse and entertain the infant while using compressed and less variable contours to soothe and reduce arousal (Falk, 2011; Papoušek, Papoušek, & Symmes, 1991). At the same time, pitch variability is not random as it is constrained by a limited repertoire of meaningful contours (Fernald, 1989). Thus, infants can still benefit from higher arousal levels through enhanced variability while simultaneously identifying and extracting the well-known melodic contour. Similarly, greater nested temporal clustering may highlight prominence contrasts and movement over time in the acoustics of ID registers. More contrastive temporal patterns could help infants both to stay tuned to the interaction and to enhance the salience or discriminability of the hierarchical structure of these patterns (de Diego-Balaguer, Martinez-Alvarez, & Pons, 2016; Delavenne, Gratier, & Devouche, 2013). Future research could investigate in more detail how nested temporal clustering relates to nested linguistic units and whether it may help infants discover and building hierarchical representations of language (e.g., via mechanisms of auditory grouping or statistical learning, cf. Echols, 1993; Hawthorne, Mazuka, & Gerken, 2015).

First indications for a *social* function of nested temporal clustering come from a recent study of Abney, Warlaumont, Oller, Wallot, and Kello (2016) demonstrating that, in naturalistic interactions, adults adapt the amount of nested clustering in their speech to the

Table 1
Utterance-final predictors of AF slope differences between AD and ID story.

	B	SE	t	p
Constant	0.016	0.010	1.606	0.143
Syllable duration (CV)	0.293	0.063	4.632	0.001**
Word duration (CV)	−0.369	0.067	−5.516	<0.001**
Phrasal constituent duration (CV)	0.438	0.070	6.299	<0.001**
PFL (SD)	0.140	0.029	4.747	0.001**
Speech rate (syll/s)	−0.076	0.013	−5.830	<0.001**

Model: *Adjusted* $R^2 = 0.936$, $F(5,9) = 42.17$, $p < 0.001$, BIC: −69.36.

clustering found in infant vocalizations. Further evidence points in the same direction showing that nested temporal clustering varies with different social and communicative contexts. In AD speech, Abney et al. (2014) found that the degree of nested clustering of acoustic onset events distinguished between two different types of adult conversations. In particular, contentious conversations with stricter turn-taking also had greater nested clustering in timescales of seconds and longer compared with friendly conversations. Thus, hierarchical temporal structure may also provide cues to distinguish different mother-infant contexts (e.g., teaching vs. caregiving contexts, play vs. soothing contexts, dyadic vs. social interactive contexts, etc.; e.g. Falk, 2011; Trainor, 1996). Besides testing a potential link between arousal and higher temporal nested clustering, future research could also investigate whether nested clustering in ID registers is involved in conveying positive affect to infants who are highly attracted and reactive to happy-sounding vocal stimuli (Corbeil, Trehub, & Peretz, 2013; Kitamura & Burnham, 1998; Nakata & Trehub, 2011; Singh, Morgan, & Best, 2002).

Finally, further work is needed to investigate how nested clustering of temporal events is related to neural processing of speech signals. Recent neuroscientific findings support the basic premise that adult brains possess a dynamic, multiscale speech envelope tracking mechanism (Zoefel & Van Rullen, 2015). The mechanism is characterized as a system of hierarchically organized neural oscillations that phase-lock to nested temporal scales of the acoustic envelope in speech (mostly between 1 and 9 Hz; for a review see Morillon, Hackett, Kajikawa, & Schroeder, 2015). Gross et al. (2013) report that "temporal edges"—a type of acoustic event—support dynamic phase adjustments of neural oscillations, and increased hierarchical coupling across frequency bands of neural oscillations. The authors propose that this increased coupling could reflect a process that aligns phases of neural excitability to salient events in speech to help process corresponding linguistic information. Little is known about the development of these neural mechanisms in infancy, or whether they are preferentially engaged by nested clustering in ID speech. Research along these lines may help to illuminate neural mechanisms that underlie sequential and relational processing during language acquisition (Gervain, Berent, & Werker, 2012).

Acknowledgements

The research leading to these results has received funding from AIRS, a MCRI-SSHRC Canada (www.airsplace.ca) research grant as well as from the European Union Seventh Framework Program [FP7/2007-2013; FP7-PEOPLE-2012-IEF] under grant agreement no. 327586 to SF. The research was also supported by a visiting professorship at the University of Montpellier to CK. We thank Anne Zorn for help with stimulus recording. Special thanks to one anonymous Reviewer for thorough and insightful comments that were very helpful in improving the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2017.02.017.

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. http://dx.doi.org/10.1037/xge0000021.

Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2016). Multiple coordination patterns in infant and adult vocalizations. *Infancy*. http://dx.doi.org/10.1111/infa.12165.

Albin, D. D., & Echols, C. H. (1996). Stressed and word-final syllables in infant-directed speech. *Infant Behavior and Development, 19*, 401–418. http://dx.doi.org/10.1016/S0163-6383(96)90002-8.

Allan, D. W. (1966). Statistics of atomic frequency standards. *IEEE Proceedings, 54*, 221–230.

Aslin, R., Woodward, J., LaMendola, N., & Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In J. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Mahwah, NJ: Lawrence Erlbaum Associates.

Aylett, M. P., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31–56. http://dx.doi.org/10.1177/00238309040470010201.

Bernstein-Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics, 14*, 303–309.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341–345.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584–1595. http://dx.doi.org/10.1111/j.1467-8624.1990.tb02885.x.

Corbeil, M., Trehub, S. E., & Peretz, I. (2013). Speech vs. singing: Infants choose happier sounds. *Frontiers in Psychology, 4*, 372. http://dx.doi.org/10.3389/fpsyg.2013.00372.

Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. *Proceedings of Speech Prosody 2002, Aix en Provence* (pp. 121–126).

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology, 7*, 44. http://dx.doi.org/10.3389/fpsyg.2016.00044.

Delavenne, A., Gratier, M., & Devouche, E. (2013). Expressive timing in infant-directed singing between 3 and 6 months. *Infant Behavior & Development, 36*(1), 1–13. http://dx.doi.org/10.1016/j.infbeh.2012.10.004.

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America, 97*(1), 585–592. http://dx.doi.org/10.1121/1.408825.

Eaves, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review*. http://dx.doi.org/10.1037/rev0000031.

Echols, C. H. (1993). A perceptually-based model of children's earliest productions. *Cognition, 46*(3), 245–296. http://dx.doi.org/10.1016/0010-0277(93)90012-K.

Falk, S. (2011). Melodic vs. intonational coding of communicative functions - A comparison of tonal contours in infant-directed song and speech. *Psychomusicology, 21*(1&2), 53–68.

Fernald, A. (1989). Intonation and communicative content in mothers' speech to infants: Is the melody the message? *Child Development, 60*, 1497–1510.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*, 209–221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104–113. http://dx.doi.org/10.1037/0012-1649.20.1.104.

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*(3), 477–501. http://dx.doi.org/10.1017/S0305000900010679.

Fernald, A. (2000). Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition. *Phonetica, 57*(2–4), 242–245. http://dx.doi.org/10.1159/000028477.

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience, 24*(3), 564–574. http://dx.doi.org/10.1162/jocn_a_00157.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology, 11*(12), e1001752. http://dx.doi.org/10.1371/journal.pbio.1001752.

Hawthorne, K., Mazuka, R., & Gerken, L. (2015). The acoustic salience of prosody trumps infants' acquired knowledge of language-specific prosodic patterns. *Journal of Memory and Language, 82*, 105–117. http://dx.doi.org/10.1016/j.jml.2015.03.005.

Jusczyk, P. W. (1999). Narrowing the distance to language: One step at a time. *Journal of Communication Disorders, 32*(4), 207–222. http://dx.doi.org/10.1016/S0021-9924(99)00014-3.

Kakouros, S., & Räsänen, O. (2016). Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features. *Cognitive Science, 40*(7), 1739–1774. http://dx.doi.org/10.1111/cogs.12306.

Kitamura, C., & Burnham, D. (1998). The infant's response to maternal vocal affect. *Advances in Infancy Research, 12*, 221–236.

Kuhl, P. K., Andruški, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686. http://dx.doi.org/10.1126/science.277.5326.684.

Lee, C. S., Kitamura, C., Burnham, D., & Todd, N. P. (2014). On the rhythm of infant-versus adult-directed speech in Australian English. *Journal of the Acoustical Society of America, 136*(1), 357–365. http://dx.doi.org/10.1121/1.4883479.

Leong, V., & Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLoS One, 10*(12), e0144411. http://dx.doi.org/10.1371/journal.pone.0144411.

Lowen, S. B., & Teich, M. C. (2005). *Fractal-based point processes*. New York: John Wiley.

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

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition, 156*, 52–59. http://dx.doi.org/10.1016/j.cognition.2016.07.015.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science, 26*(3), 341–347. http://dx.doi.org/10.1177/0956797614562453.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition, 129*(2), 362–378. http://dx.doi.org/10.1016/j.cognition.2013.07.015.

Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Current Opinion in Neurobiology, 31*, 230–238. http://dx.doi.org/10.1016/j.conb.2014.12.005.

Nakata, T., & Trehub, S. E. (2011). Expressive timing and dynamics in infant-directed and non-infant-directed singing. *Psychomusicology, 21*(1&2), 130–138.

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior & Development, 14*, 415–440. http://dx.doi.org/10.1016/0163-6383(91)90031-M.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development, 15*(3), 325–345. http://dx.doi.org/10.1016/0163-6383(92)80003-D.

Rao, K. S., Prasanna, S. R. M., & Yegnanarayana, B. (2007). Determination of instants of significant excitation in speech using hilbert envelope and group delay function. *IEEE Signal Processing Letters, 14*(10), 762–765.

Rhodes, T., Kello, C. T., & Kerster, B. (2014). Intrinsic and extrinsic contributions to heavy tails in visual foraging. *Visual Cognition, 22*(6), 809–842. http://dx.doi.org/10.1080/13506285.2014.918070.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. http://dx.doi.org/10.1126/science.274.5294.1926.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9*(6), 565–573. http://dx.doi.org/10.1111/j.1467-7687.2006.00534.x.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences, 108*(15), 6038–6043. http://dx.doi.org/10.1073/pnas.1017617108.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy, 3*, 365–394. http://dx.doi.org/10.1207/S15327078IN0303_5.

Sokolov, Y. N. (1963). *Perception and the conditioned reflex*. Oxford: Pergamon Press.

Stern, D. N. (1974). The goal and structure of mother-infant play. *Child and Adolescent Psychiatry, 13*(3), 402–421. http://dx.doi.org/10.1016/S0002-7138(09)61348-0.

Teich, M. C., & Lowen, S. B. (1994). Fractal patterns in auditory nerve-spike trains. *Engineering in Medicine and Biology Magazine, IEEE, 13*(2), 197–202.

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

Trainor, L. J. (1996). Infant preferences for infant-directed versus non infant-directed playsongs and lullabies. *Infant Behavior and Development, 19*, 83–92. http://dx.doi.org/10.1016/S0163-6383(96)90046-6.

Trainor, L. J., & Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics, 62*(2), 333–340. http://dx.doi.org/10.3758/BF03205553.

Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development, 20*, 383–396. http://dx.doi.org/10.1016/S0163-6383(97)90009-6.

Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: lullabies and playsongs. *Advances in Infancy Research, 12*, 43–78.

Viswanathan, G. M., Peng, C. K., Stanley, H. E., & Goldberger, A. L. (1997). Deviations from uniform power law scaling in nonstationary time series. *Physical Review E, 55*(1), 845–849. http://dx.doi.org/10.1103/PhysRevE.55.845.

Wang, Y., Seidl, A., & Cristia, A. (2015). Acoustic-phonetic differences between infant- and adult-directed speech: The role of stress and utterance position. *Journal of Child Language, 42*(4), 821–842. http://dx.doi.org/10.1017/S0305000914000439.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143–2152. http://dx.doi.org/10.1177/0956797613488145.

Zoefel, B., & Van Rullen, R. (2015). The role of high-level processes for oscillatory phase entrainment to speech sound. *Frontiers in Human Neuroscience, 9*, 651. http://dx.doi.org/10.3389/fnhum.2015.00651.