Dynamic Indicators of Mother-Infant Prosodic and Illocutionary Coordination

Eugene H. Buder, Anne S. Warlaumont, D. Kimbrough Oller, Lesya B. Chorna

School of Audiology and Speech-Language Pathology, The University of Memphis, Memphis, TN. USA

{ehbuder; awarlmnt; koller; lchorna}@memphis.edu

Abstract

This report introduces tools designed to detect and quantify ways in which caregivers and infants coordinate their face-to-face communicative interactions. The tools analyze this coordination at multiple levels, linking prosodic patterns to illocutionary aspects of prelinguistic discourse. Data include fundamental voice frequency and sound pressure level parameters extracted from recorded interactions and observers' codings of vocalizations according to their perceived illocutionary forces. In this approach, we do not assume that the infants' prosodic records associate categorically with any specific mature forms of linguistic or pragmatic constructs, but propose that the dyadic use of these parameters can be seen as evidence for the development of a foundational social system between mothers and infants upon which linguistic conventions can then be built. The tools are drawn accordingly from dynamic recurrence analysis and coupled-oscillators modeling and present possibilities for objective and quantitative indices of social interaction.

1. Introduction

During the prelinguistic period of vocal development, infants' speech parameters are chiefly in the prosodic domain. There has been much attention to prosodic aspects of speech directed at infants [1], and some research has endeavored with limited success to identify infants' acquisition of adult prosodic features [2]. It has long been observed that infants and mothers manage their interactions through prosodic parameters [3] and that vocal matching of these parameters can be observed with some regularity in very young infants [4]. Systematic and comprehensive quantification of mother-infant coordination has been developed for durational variables [5], but not for subtler patterns of pitch and amplitude variation. This report describes objective and quantitative tools suited to these prosodic parameters and provides evidence for the convergent validity of these methods as indicators of the development of interpersonal coordination in the first year of life.

2. Analyses

2.1. Materials

A mother and her daughter were digitally audio and video recorded while freely interacting for 20 minute sessions in a playroom-like laboratory setting. Each wore a wireless microphone system housed in custom clothing. These recordings were analyzed in a specialized software environment dubbed "AACT" [6] which incorporates an ActiveX implementation of the TF32 acoustic analysis program [7] for extraction of prosodic parameters (f_0 and RMS). AACT permits prosodic parameters to be extracted as statistical summaries in association with coded utterances, thus allowing for the recurrence analyses and illocutionary code analyses described below, and also outputs the parameters as time-series to allow for the synchrony analyses. Further details on the materials can be found in prior publications [e.g. 8].

2.2. Cross-Recurrence Analysis

Recurrence analysis is a graphic and quantitative approach to dynamic systems analysis which is particularly well suited to the study of synchronization of complex systems, and is especially useful for understanding effects over multiple scales with very long time series. It is also appropriate for nonlinear and nonstationary time series. Marwan et al. [9] provide a comprehensive and historical overview, and an example of an application to communicative interactions between child and caregiver at the level of syntax can be found in Dale and Spivey [10].

Cross-recurrence analysis requires two input series in which each datapoint from one series can be paired with a point from the other. Our interest in this case is particularly in the convergence of f_0 values across mother's and infant's contiguous utterances (although global cross-recurrence across all time scales is also informative). Even during face-to-face interaction, many of the vocalizations by either mother or infant occurred in continuous blocks of single-person vocalizations, i.e. strict utterance-by-utterance turn-taking was not regularly followed. To avoid the domination of the recurrence plot by these contiguous single-person vocalization, we limited our focus instead to paired mother and infant utterances. We thus ordered all utterances (overlapping or not) by time of onset, dropping all utterances that were both followed and preceded by utterances from the same speaker, and repeating utterances as needed to produce paired data. For example, if the mother's (M) and infant's (I) utterance onsets originally occurred in a sequence such as "M1-M2-I1-M3-I2-I3-I4-M5" the data submitted to the cross-recurrence plotting routine would be the f_0 means from the utterance sequence "M2-I1, I1-M3, M3-I2, I4-M5."

The essential steps followed to generate the plots displayed in Figure 1, technically known as *unthresholded (continuous) cross-recurrence plots* [9], were: (1) The two speaker's mean f_0 sequences were laid perpendicular to each other, with the mother's sequence flowing rightward along the x-axis and the infant's sequence flowing upward along the y-axis. (2) All datapoints from the mother's and infant's sequences were compared by calculating the log of the Euclidean distances between the two f_0 means (i.e., a semitone representation of the difference in Hz). For example, if the 2nd event in the mother's series had a f0 of 100 Hz and the 5th event in the infant's series had a f0 of 150 Hz., $\log(\sqrt{((100-150)^2)+1})=\log(50+1)=1.71$ would be entered at position (x=2, y=5) in the mother-infant coordinate system. (3) These distance values were plotted using light colors to represent large differences and dark shading to indicate small differences, normalized by dividing by the maximum value across all three plots. This allowed for direct comparison of the shading across plots.

Although, given differences between adults' and infants' typical f_0, these participants' f_0 values could have potentially exhibited distinctively different ranges the plots for this particular dyad revealed much convergence (and in fact, for each of the three sessions, it was observed during parameter extraction that the mother was often near or even above the infant in absolute Hz). For this reason, no normalization was applied to the input Hz data from these sessions.

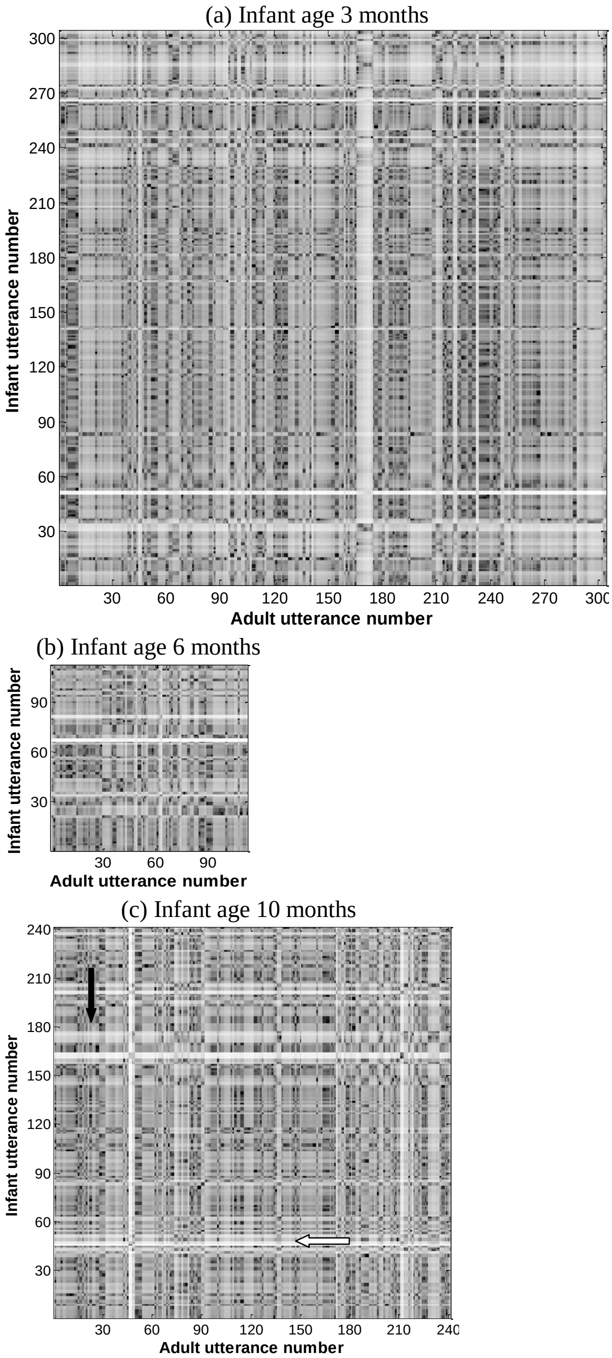

Figure 1: *Cross-recurrence plots of utterance f_0 means Dark indicates high f_0 matching and light indicates larger differences in f_0 (see text for details).*

Figure 1 displays the cross-recurrence plots from the infant AD interacting with her mother at three different ages: (a) at 4 mo., (b) at 6 mo., and (c) at 10 mo. In each plot the utterances progress in time from the origin with the adult's values progressing in sequence horizontally and the infant's values progressing in sequence vertically. The size differences between the plots reflect the numbers of contiguous utterance pairs that were available from each session (304 at 4 mo., 112 at 6 mo., and 241 at 10 mo.). With this orientation in mind, a vertical dark band suggests a series of utterances in which the mother was tending to match the infant's overall session pitch and a horizontal dark band indicates a series in which the infant was tending to match the mother's overall pitch. Lighter bands are equally interesting as indicators of pitch divergence; e.g., horizontal light bands would indicate series of infant's utterances that deviated from the mother's values.

2.3. Illocution Codes and Pitch Convergence

The infraphonological perspective [11] regards pre-linguistic infants' non-vegetative vocalizations as more or less categorically distinctive 'protophones.' Along the lines of speech act classifications first provided by Austin [12], a given instance of a protophone can be classified according to its illocutionary 'force.' For example, infants' vocalizations may be directed to their caregiver and convey complaint, exultation, or simply a desire to continue the interaction. We have developed a hierarchical and exhaustive coding scheme for such forces to be assigned to both the infant and the mother. This scheme is too elaborate to describe in detail here; what is most relevant at present is the level of distinction of person vs. object directedness. All the vocalizations in the current datasets were coded by inspection of the digital video and audio by lab personnel who were not at all oriented to the prosodic analyses and claims of this report. The codes may then be compared with indicators of pitch convergence/divergence with special attention to the infants' directedness.

The cross-recurrence plots point to episodes of vocal interaction that are distinctive in their degrees of pitch matching. These pitch convergence and divergence episodes may indicate periods of mother-infant interaction that are highly coordinated or disengaged, respectively, in the illocutionary forces of the infant's vocalizations during these periods. For example, in the 10 mo. session plotted in Figure 1(c), there is some early darker vertical banding (occurring in the 15th-20th columns of the 241 column plot: see solid black arrow). These utterance pairs are from a rhythmic game-like interaction in which the mother was naming a ball (code "Name"), and the infant was vocalizing with a social force that encouraged the game to continue (code "GameContinue"). Later, there is a distinctively lighter horizontal band (around rows 40-50, approximately 1/6 of the way up this plot: see white arrow). During this interval, the mother was attempting to continue the naming game, though the infant was socially disengaged (her illocutionary force codes during this period were exclusively "ObjectDirected").

2.4. Spectral Analysis and Sinusoidal Fitting in Rhythmic Exchange Episodes.

Further indication of the degree to which the mother and infant are prosodically coordinated can be obtained from frequency domain analyses that identify shared cyclic patterns for the participants and preferred phase relationships between these cycles. Tools for such analysis are documented with methodological details in prior reports by the 1st author [13-15] and further orientation details are provided in the caption to Figure 2 below, so only an overview of the concepts is provided here. The general goal of these analyses is to indicate and quantify degrees of "interpersonal synchrony" by adopting the concepts of coupled oscillator models [16] and some simplifying assumptions (e.g. that oscillatory patterns of prosodic data might be roughly sinusoidal in form). Based on

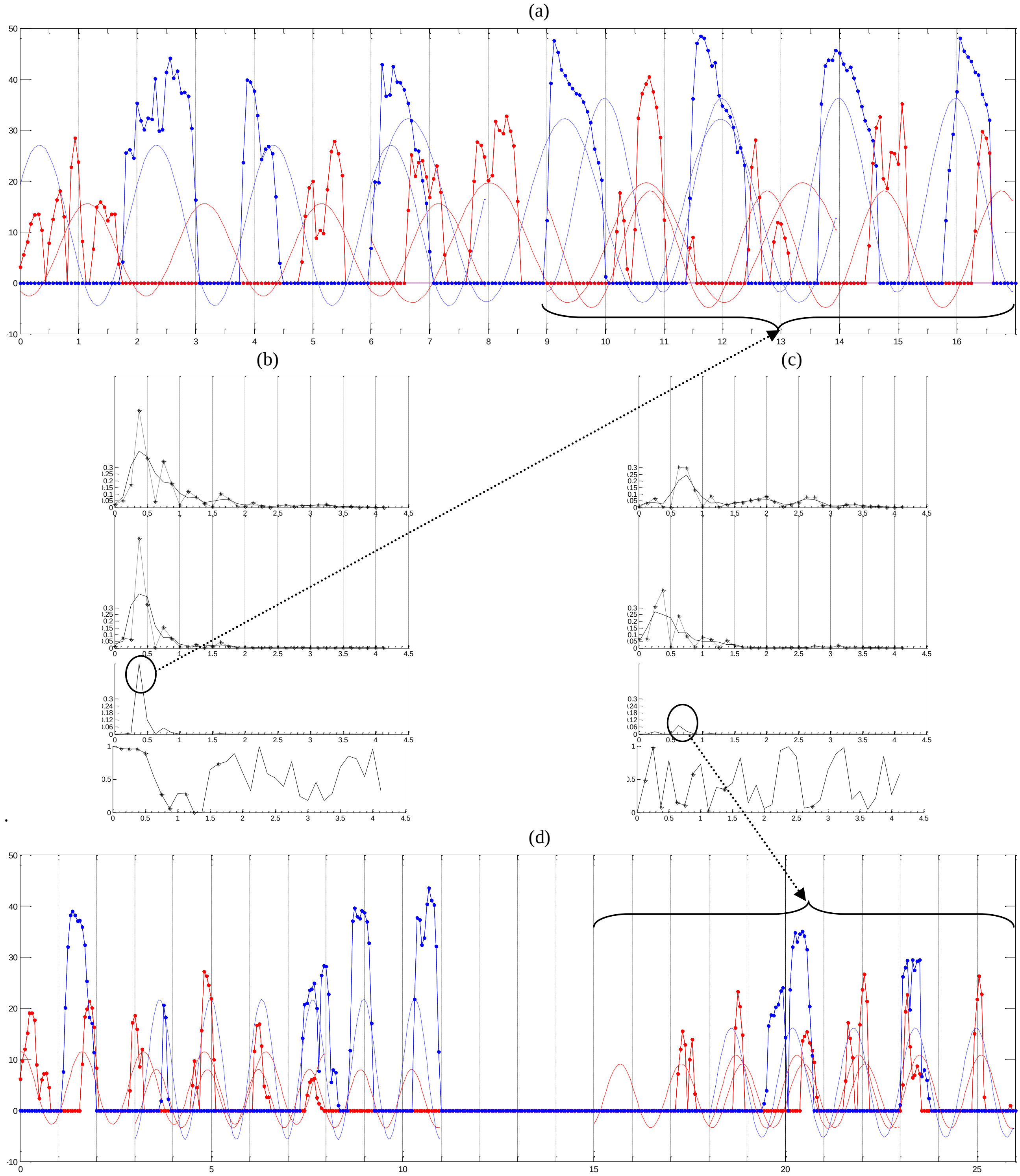

Figure 2. *Synchrony analysis plots. Panels (a) and (d) display dB-scaled RMS amplitude records from the 10-month-old infant (in solid red) and her mother (in dashed blue) for two segments of their interaction: (a) for the first 17 s episode during which the recurrence analysis indicated pitch convergence and illocution codes indicated infant's person-directedness, and (b) for the 26 s episode of pitch divergence and infant's illocutionary object-directedness. These time-domain panels also contain sinusoidal wave models indicated by frequency-domain analyses of the amplitude records, with color and line styles matching the source data. Panels (b) and (c) depict frequency domain analyses for the 8 s frames from which highest cross-spectral energy was observed in each of the segments (in both cases, as indicated by the brackets, the final 8 s). There are four analyses in each of these panels and each are aligned to the same frequency scale ranging from 0 to 4.5 Hz: the top analysis is a univariate spectrum of the infant's amplitude series for a given 8 s frame and the next section down is the univariate spectrum of the mother's amplitude data for that frame (in each, the raw FFT output, or periodogram, is drawn with a dashed line, and a 3-point smoothing of this result is drawn with a smoothed line to provide an estimate of the power spectrum). The third section is the unsmoothed bivariate cross-spectrum. The cross-spectral peaks used to identify specific wave models are circled. The fourth and lowest section is a bivariate phase spectrum in which individual phase results are transformed to a 0 to 1 scale indicating the phase relationship, with 1 at the top representing an anti-phase or 180° relationship and 0 at the bottom representing an in-phase or 0° relationship.*

prior analyses, and in accord with the framework of "coordination dynamics" [17], it is presumed that stable phase relationships between cycles with similar frequencies are evidence of coordination or synchrony, and that in-phase (0°) or anti-phase (180°) relationships are particularly preferred.

As seen in the relatively peaked form of the univariate frequency-domain results in the upper two plots of panels (b) and (c) in Figure 2, spectral analyses of the amplitude contours of the mother's and infant's vocalizations do indicate dominant cycles. Of particular interest here is the comparison of the two segments from the session for the infant at 10 months of age, as identified above by inspection of the cross-recurrence plots and illocutionary codes, for patterns that indicate varying degrees of dyadic level coordination.

Both participants' amplitude contours from both segment could be modeled with simple sinusoidal contours based on univariate (single speaker) spectra, but interpersonal coordination was clearly stronger in the first segment as seen in panels (a) and (c). In particular, the cross-spectral magnitude was much stronger in this segment (c. 6.0) in comparison to the other (< 0.6), reflecting the fact that the clearly dominant cyclicity was the same for both participants, at 0.375 Hz, but at different cycle lengths in the less interactive segment as seen in the univariate spectra of panel (c). Furthermore, the phase spectrum for the interactive segment, in the lowest section of panel (b), showed a clear anti-phase locking for the modeled cycle and other lower frequencies as well. Because the modeling of cyclic structure in the first (pitch convergent/person-directed) session could be based in this example on the cross-spectral peak magnitude, exactly the same modeling strategy could be applied to the second (pitch divergent/object-directed) session to yield the sine-wave overlays seen in panel (d). Although these wave models do appear to capture some degree of rhythmic structure, several aspects of the results display much less interpersonal synchrony than that observed in the first session: the mother's and infant's dominant cycles as seen in the univariate spectra of panel (c) are clearly at different frequencies; as noted before the cross-spectral magnitude is much reduced; and to the extent that a phase relationship is clear at the selected frequency of the final frame (0.625 Hz), it is in an in-phase relationship. The in-phase relationship is interpretable as representing the mother's attempt to re-engage the infant (especially after the extended mutual silence in the center of this episode) through simultaneous locutions.

3. Discussion and Conclusions

The specific segments analyzed in this report were selected only according to inspection of the f_0 cross-recurrence plots and it so happened that they correlated with illocutionary directedness and differing degrees of dyadic synchrony. The recurrence plot provides a global overview of pitch convergence/divergence and can yield many quantitative indices not explicated here [9]. Synchrony analyses provide more locally focused views of tendencies for cycles in prosodic parameters to be coordinated across participants, and although only cycles of amplitude were illustrated here, the techniques are equally applicable to cycles of f_0 (which were also evident in the current data especially in the more pitch convergent person-directed segment). We also suspect that prosodic coordination *per se* is probably of incidental or even epiphenomenal value unless reflected in other levels of communication analysis like illocutionary force. The converging indication of coordination across the three analysis approaches in the segments analyzed here suggests the potential for these tools for identifying and quantifying meaningful prosodic coordination between infant and caregiver. Further application of these tools to the remainder of the corpus will clarify the extent of such cross-level validation.

Finally, utilizing quantitative indices that can be derived from each of the three analysis tools demonstrated here, our ongoing goal is to quantify developmental and comparative effects in this and other mother-infant interaction corpora.

4. Acknowledgements

Support for this work was provided by NIH/NIDCD grant DC006099 awarded to the 1st and 3rd authors and a Department of Energy Computational Science Graduate Fellowship awarded to the 2nd author. The authors also thank Rick Dale, PhD, for assistance with recurrence analysis computations and Heather Ramsdell, PhD for the illocutionary coding and other general assistance.

5. References

[1] A. Fernald, "Intonation and communicative intent in mothers'speech to infants: Is the melody the message?," *Child Development,* vol. 60, pp. 1497-1510, 1989.

[2] D. Snow and H. Balog, "Do children produce the melody before the words? A review of developmental intonation research," *Lingua,* vol. 112, pp. 1025-1058, 2002.

[3] D. N. Stern, "Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors," in *The effect of the infant on its caregiver*, L. A. Rosenblum, Ed. New York: Wiley, 1974, pp. 187-213.

[4] M. Papoušek and H. Papoušek, "Forms and functions of vocal matching in interactions between mothers and their precanonical infants," *First Language,* vol. 9, pp. 137-158, 1989.

[5] J. Jaffe, B. Beebe, S. Feldstein, C. L. Crown, and M. D. Jasnow, "Rhythms of dialogue in infancy: Coordinated timing in development," *Monographs of the Society for Research in Child Development,* vol. 66, 2001.

[6] R. E. Delgado, "AACT - Action Analysis Coding and Training Software," Miami, FL: Intelligent Hearing Systems Corp., 2008.

[7] P. Milenkovic, "TF32," Madison, WI: University of Wisconsin-Madison, 2001 [URL: http://userpages.chorus.net/cspeech/].

[8] E. H. Buder, L. Chorna, K. Oller, and R. Robinson, "Vibratory regime classification of infant phonation," *Journal of Voice,* vol. 22, pp. 553-564, 2008.

[9] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports,* vol. 438, pp. 237-329, 2007.

[10] R. Dale and M. J. Spivey, "Unraveling the dyad: using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation," *Language Learning,* vol. 56, pp. 391-430, 2006.

[11] D. K. Oller, *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[12] J. L. Austin, *How to do things with words*. London: Oxford Univ. Press, 1962.

[13] E. H. Buder and J. L. Edrington, "Conversational prosodic interactivity when one partner has aphasia," in *Speech Prosody 2008 Conference*, P. A. Barbosa, S. Madureira, and C. Reis, Eds. Campinas, Brazil: Editora RG/CNPq, 2008, pp. 501-504.

[14] E. H. Buder and A. Eriksson, "Prosodic cycles and interpersonal synchrony in American English and Swedish," in *Eurospeech 1997 Proceedings*. vol. 1, E. Dermatas, Ed. Grenoble, France: European Speech Communication Association, 1997, pp. 235-238.

[15] E. H. Buder and A. Eriksson, "Time-series analysis of conversational prosody for the identification of rhythmic units," in *Proceedings of the 14th International Congress of Phonetic Sciences*. vol. 2, A. C. Baile, Ed. San Francisco, CA: University of California, Berkley, 1999, pp. 1071-1074.

[16] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A universal concept in nonlinear sciences*. Cambridge, UK: Cambridge University Press, 2001.

[17] J. A. S. Kelso, *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: The MIT Press, 1997.