PROSODIC CYCLES AND INTERPERSONAL SYNCHRONY IN AMERICAN ENGLISH AND SWEDISH

Eugene H. Buder

School of Audiology & Speech-Language Pathology
University of Memphis
807 Jefferson Ave., Memphis, TN 38105 USA
ehbuder@cc.memphis.edu

Anders Eriksson

Department of Phonetics
Umeå University
S-901 87 Umeå, Sweden
anderse@ling.umu.se

ABSTRACT

The paper addresses the question of rhythmic structuring of conversational interaction. Conversational speech requires active co-operation and co-ordination of the behavior of two or more speakers. Previous research indicates that one of the mechanisms used by speakers to regulate conversational interaction, is close monitoring and adaptation to rhythmic patterns. When this does not function properly, interaction may be adversely affected or even break down. There are reasons to believe that these mechanisms are used universally across languages, but there are also likely to be patterns that are language-specific. The research project, of which the present paper forms a first published report, is an attempt at separating the universal and language-specific aspects of the regulating rhythmic patterns. Although this research is primarily meant to clarify the mechanisms of conversational interaction from a linguistic/phonetic point of view, its applicability to speech technology is evident. Growing interest in dialogue systems for applications to man-machine communication demands more detailed data on all aspects of natural human conversation.

1. INTRODUCTION

Conversational communication functions by the simultaneous engagement of processes of speech production and perception as people interleave the acts of listening and speaking. Conversational speech by one person must be patterned so as to somehow "fit" the patterns of the other people engaged in that conversation. Effective turn-taking is one example of this fit, but there are other ways in which speech is temporally structured during conversational interaction. Growing interest in dialogue systems for applications to man-machine communication demands data on natural human conversation, and the rhythmic prosodic structure of conversation has emerged as one important mechanism by which conversational speech is managed.

2. BACKGROUND

It is clear that speech is acoustically patterned in ways that produce rhythmic features (e.g., [4], [7]), though the precise physical nature and functions of rhythmic structures are not always so clear. Some social scientists have begun to explore the possibility that these structures help to organize social interaction (e.g., [1], [8], [9]). Previous research by Buder ([1], [2], [3]) has produced examples of rhythmic patterning in acoustic speech parameters of fundamental voice frequency and intensity (pitch and loudness) that appear to explain the timing of speech behaviors in both partners in dyadic (two-person) conversations. Buder's work indicates that statistical models based on time-series analysis are adequate for detecting such patterning.

This approach suggests that there may be temporal structures in speech behavior that are independent of specific linguistic units. For example, units such as phonemic clause [6] or breathgroup [5], usually studied in isolated speech samples, may function at the conversational level of analysis and be controlled partly by social interaction demands. As the demands of social interaction may be considered to be linguistically "universal," it is probable that these mechanisms are to be found across all the languages of the world. It is also possible that the ways in which rhythmic structures are employed as organizing mechanisms for social interaction may vary depending on the ways particular languages constrain the production of rhythmic units, as there are known to be many differences in such prosodic elements from one language to another.

3. METHOD AND TYPICAL RESULTS

Recordings of 15-minute conversations were made in the US and in Sweden using both normal lavaliere-type microphones attached to the subjects' clothing and contact-type (accelerometric) microphones adhered to the subjects' throat. Digitized signals from both microphones have been acquired from a half-minute long segment approximately 12 minutes into four of these conversations (one between men and one between women from each language) and analyzed for fundamental frequency (f0) and intensity using the CSpeechSP acoustic analysis program. The contact microphones optimize fundamental frequency analysis and provide signals with an absence of "cross-talk," while the normal microphones provide the signal for vocal intensity analysis. Fundamental frequency traces

were hand-checked to remove artifacts due to aberrant voicing or algorithm failure, and the intensity traces were edited to remove cross-talk (using the contact microphone signals as a guide where necessary, especially during simultaneous speech by both participants). The acoustic parameter data were then exported from the analysis environment at a rate of 240 samples per second, and then further downsampled (using median smoothing for the f0 data and mean smoothing for the intensity data) in successive 5- and 3-point smoothings to obtain time-series data at 16 samples per second.

Data such as these were then further transformed for spectral analysis. The essential steps similar to both parameters were as follows 1) framing of the data into 128 point frames with 48 point overlaps, 2) normalization of the data within a frame to the overall sample SD and to the within-frame mean, 3) detrending and cosine edge tapering, and 4) transformation by FFT into periodogram form, and 5) 3-point smoothing of the periodogram. An additional step was required for the f0 data due to the presence of gaps in the data; these gaps were zero-padded within each frame after the mean normalization. In previous research with lower sampling rates ([1], [2]), cross-spectra were also obtained to assess shared cyclic activity; here we proceeded on the basis of univariate spectra.

Current results were summarized primarily in graphic form with sinusoidal models based on the spectral analyses superimposed on the prosodic data to assess the interpretive significance of the cycles. Figure 1 displays one such interpretation applied to 23 seconds of conversation between two Swedish women. Two sets of panels in the center of this figure display the prosodic data; the display "wraps" from the upper to lower panel sets, with the first 12 seconds charted in the top set and continuing for another 11 seconds in the bottom set. In each set, the top panel contains f0 data and the bottom panel contains intensity data. One speaker's data (person "A") is displayed using filled circles, and the other (person "B") with hollow circles. In the intensity panel the circles are connected with solid and dashed lines, respectively, and the f0 data are left unconnected. Note that person A's speech dominates the top panels, with a turn-exchange that occurs sometime after second 10, yielding a predominance of B's speech in the lower panels. As indicated by the text inserts and arrows, sinusoidal models superimposed on the prosodic data were based on graphic inspection of selected frame's spectra and the frequency and phase information determined by these spectra. For the four models displayed here, the associated spectra are displayed with periodogram points connected by dashed lines and the smoothed spectra connected by solid lines. (Continuing the convention established in the time-domain data, person A's periodograms are displayed with filled circles and person B's with hollow circles).

The spectral results and the sinusoidal models present a compelling picture of prosodic cycles in these data, and we believe that these cycles play a role in the organization of conversational interaction. For example, the clearly significant perodicity of person A's intensity data as mapped in the top set of panels is extended and echoed by person B across the turn-exchange. This is illustrated in Figure 1 by the extension of the solid line of A's intensity model by a dashed line extrapolation that extends to approximately second 15. This extrapolated cycle is seen to align nicely in phase with the independently determined dotted line model of B's intensity data. This demonstrates what is in our experience a relatively common occurrence; the continuance of speakers' "rhythms" across turn-exchanges. A similar phenomenon can be observed in the f0 data, in which we see that the frequency of greatest magnitude in A's data can also be observed as a cycle in B's data. Although this is not the cycle with greatest magnitude in B's spectral analysis, it can be seen that the cycle is clearly aligned with virtually all of the dominant pitch peaks in her data.

As illustrated in the accompanying poster, all four conversational samples we have investigated for this report (men and women from both language communities) contain strong spectral cycles, and sinusoidal models based on these cycles were found to align meaningfully with conversation structure. In three of the four extracts (i.e., excluding the American male-male sample) the chief (highest magnitude) cycles identified by the spectral analyses of at least one parameter (usually vocal intensity, but also in sections of f0) occurred in clear phase alignment across both partners. In other words, the period and phase of a prosodic cycle initiated by one partner was maintained by the other, across speaker transitions, in three of the four 23-second samples we have investigated to date. Typical cycle wavelengths detected by the technique have been on the order of 1 to 4 s for intensity cycles, and 0.6 to 2.5 s for f0 cycles. We are also intrigued by an apparent convergence between these f0 cycles and similar "quantal" tendencies described by Fant and Kruckenberg for segmental durations in read Swedish [7].

5. CONCLUSIONS

Analyses of data suggest that in conversations between same-gender strangers, the patterns of rhythmic integration, or "synchrony," are quite similar in both American English and Swedish.

Although this tentatively supports the hypothesis that interaction related speech rhythms are universal to both languages, more extensive analyses are underway to further validate these findings. Via analysis of the alignment of these cycles with linguistic units and therefore language-specific prosodic elements, we plan to assess the possibility that differences between these languages influence the interaction patterns, though it also remains possible that the interaction patterns are in
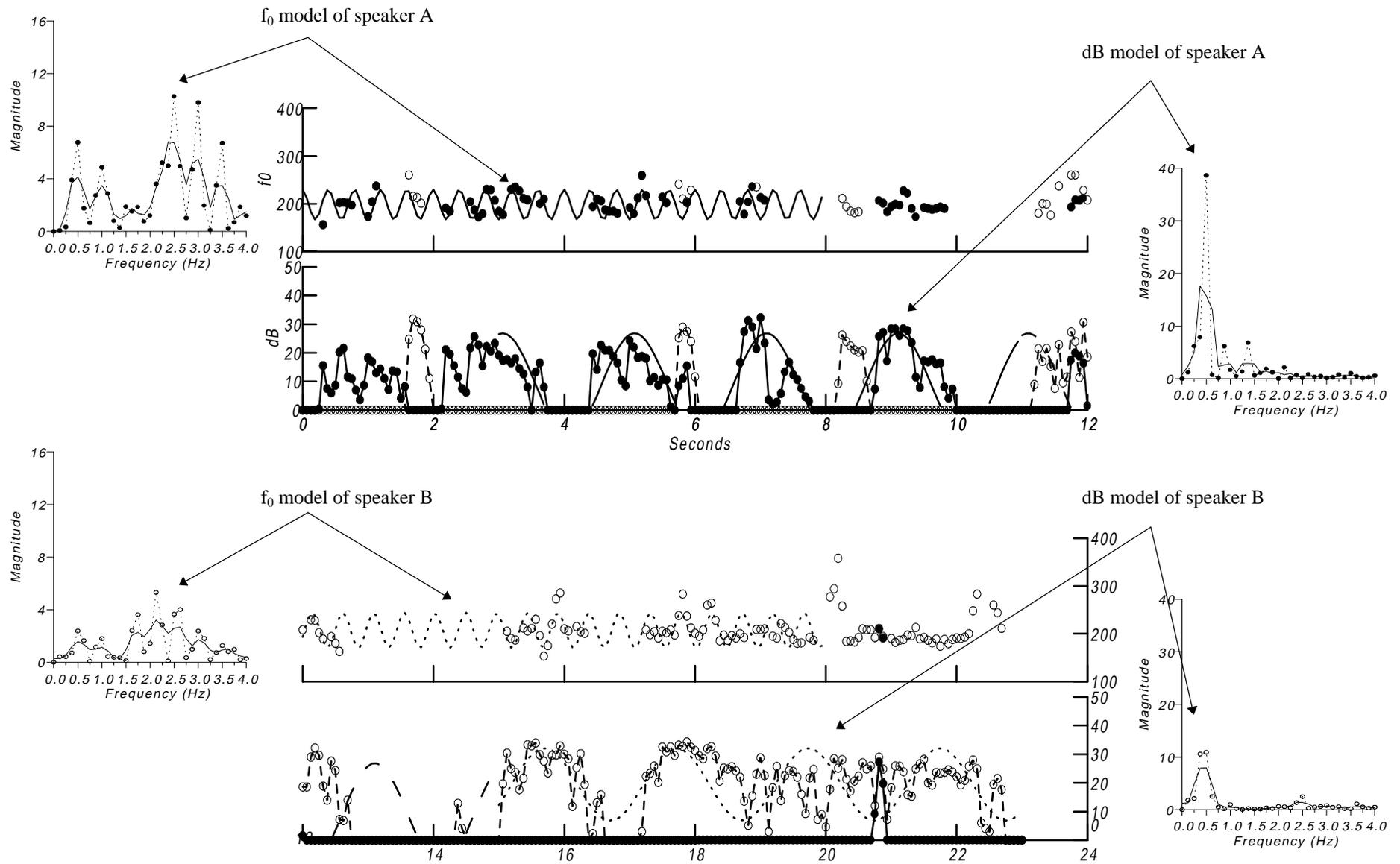
Figure 1. Prosodic data from a passage of conversation between two Swedish women, with spectral analyses and cyclic models of selected portions (see text for further explanation).

fact preeminent and bear no specific relation to the phonological structures of the languages in which they appear. The results of this research may therefore ultimately address basic issues of language and culture, but we also anticipate the possibility that interpersonal synchrony patterns adhere to universal principles of dialogue management.

6. REFERENCES

[1] Buder, E. H. (1996) Dynamics of speech processes in dyadic interaction. In J.H. Watt & C. A. VanLear, (Eds.), *Dynamic Patterns in Communication Processes* (pp. 301–325). Thousand Oaks, CA: Sage.

[2] Buder, E.H. (1991) *Vocal synchrony in conversations: spectral analysis of fundamental voice frequency*. Unpublished doctoral dissertation, Departments of Communication Arts and Communicative Disorders, University of Wisconsin, Madison, WI.

[3] Buder, E. H. (1986). Coherence of speech rhythms in conversations: Autocorrelation analysis of fundamental voice frequency. *Toronto Semiotic Circle Monograph*, (1986)2, Toronto: Toronto Semiotic Circle.

[4] Eriksson, Anders. (1991) *Aspects of Swedish Speech Rhythm.* Göteborg: Gothenburg Monographs in Linguistics 9.

[5] Lieberman, P. (1967) *Intonation, Perception, and Language*. Cambridge, Mass.: M.I.T.Press.

[6] Trager, G. L., & Smith, H. L., Jr. (1951/1965) *An outline of English structure*. New York: American Council of Learned Societies.

[7] Fant, G., & Kruckenberg, A. (1996). On the quantal nature of speech timing. In *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 2044–2048).

[8] Warner, R.M.. (1996). Coordinated cycles in behavior and physiology during face-to face social interactions. In J.H. Watt & C.A. VanLear (Eds.), *Dynamic patterns in communication processes* (pp. 327–352). Thousand Oaks, CA: Sage.

[9] Cappella, J.N. (1996). Dynamic coordination of vocal and kinesic behavior in dyadic interaction: methods, problems, and interpersonal outcomes. In J.H. Watt & C.A. VanLear (Eds.), *Dynamic patterns in communication processes.* (pp. 353–386). Thousand Oaks, CA: Sage.