

Finding acoustic evidence for speech rhythm across languages : A speech cycling approach

Keiichi Tajima,**** Bushra A. Zawaydeh,***** Mafuyu Kitahara,** and Robert F. Port***

* Information Sciences Division, ATR International,
2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan
E-mail : ktajima@isd.atr.co.jp

** Lernout & Hauspie Speech Products USA,
52, Third Avenue, Burlington, MA 01803, USA
E-mail : bzawaydeh@lhs.com

*** Department of Linguistics, Indiana University,
322, Memorial Hall, Indiana University, Bloomington, IN 47405, USA
E-mail : mkitahar@indiana.edu (Kitahara) ; port@indiana.edu (Port)
(Received 27 April 2000)

Keywords : Speech rhythm, Stress timing, Mora timing, Phase angle, Beats.

PACS number : 43.70.Fq

1. Introduction

Many acoustic-phonetic studies have attempted to find physical correlates of the common impression that languages are spoken with different kinds of rhythm. However, it has proven extremely difficult to find evidence for these impressions. For example, a comparison of five languages (Dauer, 1983) showed that so-called "stress-timed" languages such as English do not show a stronger tendency for stressed syllables to occur at equal time intervals than do so-called "syllable-timed" languages such as Italian. Some researchers have indeed claimed that speech rhythm is an illusory phenomenon attributable primarily to perceptual mechanisms (Lehiste, 1977).

This paper describes a novel experimental paradigm in which acoustic evidence for speech rhythm is sought in utterances produced in a purposefully rhythmic context. It is hypothesized that speech rhythm is empirically observable as *preferred patterns of temporal organization* under suitably constrained speaking conditions. This approach was inspired by studies in human motor behavior, such as the "finger wagging" task (Kelso, Southard, and Goodman, 1979), in which subjects oscillate the index fingers of both hands back and forth in time with a metronome. Under changes in the rate of oscillation of the fingers, it was found that subjects could comfortably oscillate the fingers in only a small number of ways, namely, either an in-phase pattern in which the fingers move toward and away from each other at the same time, or an anti-phase pattern, in which they move to the left and right at the same time.

Applying this method to speech led to the "speech cycling" task, in which subjects repeat whole phrases in time with periodic auditory stimuli. Previous speech cycling studies (Cummins and Port, 1998 ; Tajima, 1998) have found results similar to those of simple

motor tasks ; under repetitive contexts, there are only a small number of stable timing patterns in which a phrase can be produced. In this paper, speech cycling data from three languages are compared. Arabic and English have been labeled as "stress-timed" (Roach, 1982), and Japanese as "mora-timed" (Port, Dalby, and O'Dell, 1987). Given these labels, this task should reveal similarities between Arabic and English, and qualitatively different results for Japanese.

2. The speech cycling task

Four native speakers each of Ammani-Jordanian Arabic, Tokyo Japanese, and American English participated. On each trial, subjects saw a phrase on the computer screen, and heard an isochronous series of 600-Hz, 50-ms tones presented through headphones. They listened to the first four metronome beeps, then joined in to repeat the phrase, aligning the beginning of the phrase with each successive beep. They repeated the phrase eight times in a single breath, at which point the metronome stopped. The metronome period was initially set to about 1,500 ms, and decreased by 7% on successive trials. When the metronome became too fast, the period was reset and the subject proceeded to the next phrase. On average, subjects repeated each phrase at about 7-10 different rates.

3. Measurement and analysis

A "beat extractor" (Cummins and Port, 1998) was used to convert subjects' productions into a series of "beats" placed near the vowel onsets of syllables. First, a band-pass filter with 1,000-Hz center frequency and 600-Hz bandwidth was applied to the speech signal to retain primarily the formant energies. Then, the signal was rectified and low-pass filtered at 20 Hz to yield a smooth energy profile (see Fig. 1). Finally, for every local peak in the profile that exceeded a given

threshold, a “beat” was placed exactly halfway between points in time where the local amplitude rise was 10% complete and 90% complete. The program output was corrected by hand. For syllables with nasal or voiced stop consonants, beats located in this fashion occur close to the so-called “perceptual moment of occurrence”, or P-center, of the syllable (Scott, 1993).

Timing pattern of the repetitions was analyzed by obtaining the *phase angle* of the phrase-final syllable beat in each repetition. That is, the time of occurrence of the final syllable’s beat was expressed relative to the

cycle defined by the initial beat of the current repetition and that of the following repetition. In Fig. 1, for example, phase angle of the beat of the syllable “beer” in “buy Doug a beer” is computed as b/a .

4. Results

Phase of the phrase-final syllable was measured for repetitions at all metronome rates by all speakers in each language group. The observed phases were then collapsed into a histogram. A separate histogram was made for each test phrase. Four phrases were tested in each of the three languages, as shown in Fig. 2. In this figure, the phrases increase in length from top to bottom, from five to eight moras in Japanese, and four to seven syllables in Arabic and English. The Japanese phrases contained only one-mora syllables, so the number of moras and syllables was equivalent.

Figure 2 shows that there are noticeable distributional peaks in the histograms, with most histograms showing peaks at several phase angles. This indicates that subjects produced the phrase-final syllables at certain preferred phase angles over others. In other words, when subjects repeated a phrase at a continuum of repetition rates, they exhibited a small number of relatively stable rhythmic patterns.

Prominent distributional peaks in Fig. 2 are indicat-

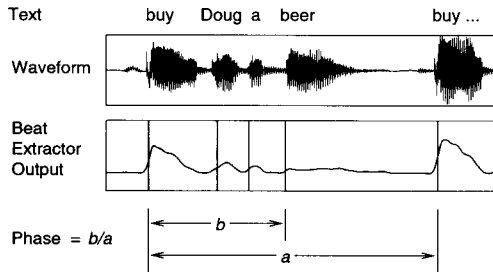


Fig. 1 Illustration of beat extraction and phase calculation in a sample waveform.

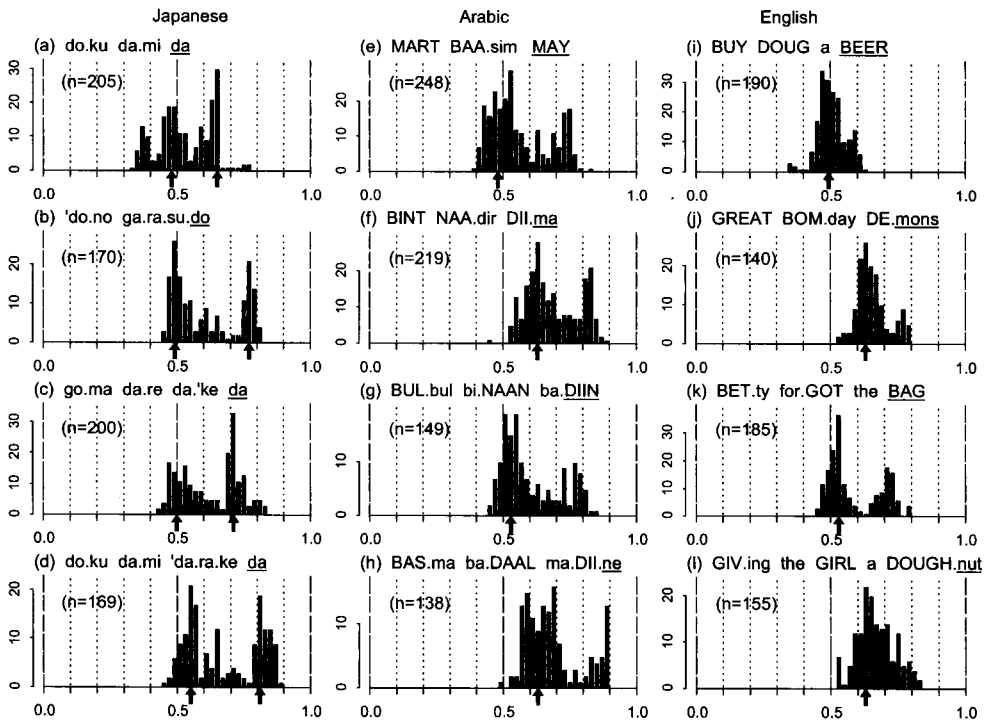


Fig. 2 Histograms of the phase of the phrase-final syllable (underlined) in Japanese, Arabic, and English phrases. Stressed syllables in Arabic and English are shown in uppercase. Pitch-accented moras in Japanese are preceded by an apostrophe (').

ed with arrows at the bottom of each histogram. When these prominent peaks are compared among phrases in each language, cross-linguistic differences in preferred rhythm become evident. In Arabic, the tallest distributional peak was close to phase 0.5 for phrases (e) and (g) in Fig. 2, but noticeably later than 0.5 for phrases (f) and (h). Similar differences were also found for English phrases (i) and (k) vs. phrases (j) and (l). Such differences in the preferred timing of the final syllable can be related to the *stress pattern* of the phrases. That is, if the final syllable was stressed (e.g., English phrase (i) "BUY DOUG a *BEER*"), then subjects preferred to produce it near phase 0.5. If it was unstressed (e.g., English phrase (j) "GREAT BOM. bay *DE.mons*"), however, it was produced at a later phase than 0.5; further examination of the data revealed that the preceding *stressed* syllable, rather than the phrase-final syllable, was produced near phase 0.5. Altogether, then, English and Arabic speakers showed a preference to produce the last stressed syllable of the phrase near phase 0.5, *i.e.*, halfway between the start of successive repetitions.

Histograms for the Japanese phrases seem to show a qualitatively different result. A distributional peak was found close to phase 0.5 for all phrases. It appears that Japanese speakers found it comfortable to repeat each phrase in such a way that the phrase-final mora was produced near phase 0.5, *i.e.*, halfway between the start of successive repetitions.

The differences between Japanese and the other two languages may have been obtained because the Japanese phrases did not show the specific stress pattern variation that the Arabic and English phrases did. Nevertheless, Fig. 2 reveals an additional aspect of rhythmic preference in Japanese that seem to be qualitatively different from Arabic and English. Each Japanese phrase in Fig. 2 showed a second prominent histogram peak at a phase later than 0.5, *e.g.*, near 0.65 for phrase (a). The peak in (a), however, was earlier in phase than that in (b) which occurred near 0.8; similarly, the peak was earlier in (c) than in (d). These relative differences can be related to the length of the phrase; phrases with an *even* number of moras—(b) and (d)—were generally produced with shorter intervals between successive repetitions than phrases with an *odd* number of moras—(a) and (c). It appears that native Japanese speakers preferred to insert a short rest between repetitions for phrases with an odd number of moras, suggesting the potential influence of units that are two moras in length in the timing of Japanese

utterances (Sakano, 1996).

5. Conclusion

Data in this paper provide observable correlates of conventional descriptions about rhythmic similarities and differences across languages. Arabic and English speakers seemed to pay particular attention to the stressed syllables of the phrases, producing the final stress near phase 0.5. In this sense, both languages are "stress-timed", and both show qualitative differences from Japanese, a "mora-timed" language. Japanese speakers sometimes produced the phrase-final mora near phase 0.5, and sometimes showed distinct patterns depending on the number of moras in the phrase.

Even though the above results are not likely to be attested in ordinary spoken language, the data suggest that speech rhythm is not entirely a perceptual illusion. Rather, rhythmic regularity becomes observable in speech production as patterns that are preferentially produced by native speakers under suitably constrained conditions. Stable rhythmic behavior that arises in these constrained conditions can be profitably applied in the investigation of language-specific and language-general aspects of speech rhythm.

References

- Cummins, F. and Port, R. F. (1998). "Rhythmic constraints on stress timing in English," *J. Phonet.* **26**, 145–171.
- Dauer, R. M. (1983). "Stress-timing and syllable-timing reanalyzed," *J. Phonet.* **11**, 51–62.
- Kelso, J. A. S., Southard, D., and Goodman, D. (1979). "On the nature of human interlimb coordination," *Science* **203**, 1029–1031.
- Lehiste, I. (1977). "Isochrony reconsidered," *J. Phonet.* **5**, 253–263.
- Port, R. F., Dalby, J., and O'Dell, M. (1987). "Evidence for mora timing in Japanese," *J. Acoust. Soc. Am.* **81**, 1574–1564.
- Roach, P. (1982). "On the distinction between 'stress-timed' and 'syllable-timed' languages," in *Linguistic Controversies*, Crystal, D., Ed. (Edward Arnold, London), pp. 73–79.
- Sakano, N. (1996). *Shichi-go-chō no Nazo wo Toku (Solving the Mystery of the Seven-Five Meter)* (Taishūkan Shoten, Tokyo) (in Japanese).
- Scott, S. K. (1993). "P-centres in speech: An acoustic analysis," *Doct. Diss.*, Univ. Coll. London.
- Tajima, K. (1998). "Speech rhythm in english and japanese: Experiments in speech cycling," *Doct. Diss.*, Indiana Univ., Bloomington, IN, USA.