A COMPARATIVE STUDY OF SPEECH RHYTHM IN ARABIC, ENGLISH, AND JAPANESE

Keiichi Tajima*, Bushra, A. Zawaydeh*, and Mafuyu Kitahara*
* Indiana University, Bloomington, IN, USA
† ATR Human Information Processing Research Laboratories, Kyoto, Japan

ABSTRACT
Conventional rhythmic distinctions such as "stress-timing" vs. "syllable-timing" were investigated using a novel technique whereby speakers produced a phrase repeatedly along with a metronome. When stable rhythmic productions of phrases in Arabic, English, and Japanese were compared, Japanese speakers preferred to place the phrase-final syllable at "simple harmonic phases", such as halfway between the start of successive repetitions. Arabic and English speakers, by contrast, placed stressed syllables at these points. However, Arabic and English were found to differ from each other in the timing of stressed syllables within a phrase. English speakers more closely approximated isochrony of stressed syllables than did Arabic speakers.

1. INTRODUCTION
Abercrombie [1] claimed that languages fall into rhythmic categories such as "stress-timed" and "syllable-timed". Arabic and English have been classified as "stress-timed", and stress plays a crucial role in the phonological system of both languages. This contrasts with Japanese—a "mora-timed" language [2]—where stress plays no role in the global organization of utterances. However, it has proven difficult to find acoustic evidence that support these typological distinctions [3]. Units such as stressed syllables or individual moras do not occur at regular time intervals.

Recently, an experimental method called "speech cycling" has been developed to study speech rhythm [4,5]. Subjects produce a phrase repeatedly in time with a periodic auditory stimulus. When speakers repeat a phrase at a continuum of increasing speaking rates, they have been found to fall in and out of stable, quasi-discrete rhythmic modes [5]. In this paper, we compare the results of speech cycling experiments in three languages: Arabic, English, and Japanese.

2. METHODS
2.1. Speakers
Data were collected from four native speakers each of Amman Jordanian Arabic, American English, and Tokyo Japanese.

2.2. Text materials
There were 6 Arabic, 4 English, and 4 Japanese phrases. The Japanese phrases, shown in Table 1, were 5-8 moras in length; all syllables were monomoraic CV syllables. The Arabic and English phrases, shown in Table 2, were 4-7 syllables in length. The pattern of stressed (s) and unstressed (u) syllables varied across phrases, but was matched between the two languages as shown in the table. (The phrases do not match completely because the Arabic and English experiments were originally conducted as separate studies.) Each phrase contained 3 stressed syllables. The first and second stressed syllables were intervened by varying numbers of unstressed syllables. The third stressed syllable was either the final or penultimate syllable of the phrase.

2.3. Instructions and procedure
Subjects listened to an isochronous series of 50-ms, 600-Hz tones through headphones. On each trial, 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 produced the phrase eight times without inserting breaths between repetitions.

For each phrase, the metronome was initially set to a slow rate, with a period of about 1500 ms. On each successive trial, the metronome period was decreased to 93% of the previous trial. When the rate became too fast for the speaker to repeat the phrase accurately, then it was reset to a slow rate, and the subject proceeded to the next phrase. On average, subjects repeated each phrase at 7-10 different metronome rates.

2.4. Measurement and analysis
The speech signal was converted into a series of "beats", by placing a beat near the vowel onset of each syllable. These beats were used to capture the timing of vowel onsets in each...
Table 2: Arabic and English phrases.

<table>
<thead>
<tr>
<th>No.</th>
<th>Syll</th>
<th>Stress</th>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>ssus</td>
<td>MART</td>
<td>BAA.sim MAY</td>
<td>BUY DOUG a BEER</td>
</tr>
<tr>
<td>5</td>
<td>ssusu</td>
<td>BINT</td>
<td>NAA.dir DII.ma</td>
<td>GREAT BOM.bay DE.mons</td>
</tr>
<tr>
<td>6</td>
<td>susus</td>
<td>BA.dir BA.na BEET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>suusu</td>
<td>BUK.ra BINA.dir DAM BAA.sil</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>suusus</td>
<td>BUL.bul bi.NAAN ba.DIIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>suususu</td>
<td>BAS.ma ba.DAAL ma.DII.ne</td>
<td>CIV.ing the GIRL a DOUGH.nut</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: External and internal phases.

Figure 2: Histogram of external phase of the final syllable in the Arabic phrase “mart baasim may”.

3. RESULTS
3.1. Stable rhythmic modes
Since the metronome was made faster across trials, the speaker had to repeat each phrase at a faster rate after each trial. Thus, the interval duration between the onset of one repetition to the next (interval “a” in Figure 1) was expected to decrease successively across trials. From this, one might expect that measures such as “b/a” (i.e., external phase of a syllable), where “a” is the denominator, would get larger in a gradient fashion across trials. However, this was not the case. Figure 2 shows data from the Arabic phrase “mart baasim may”, where observed external phases of the syllable “may” are shown as a histogram. The data are collapsed across all repetitions measured, from all metronome rates done by all Arabic speakers. The figure clearly shows several distributional modes, suggesting that the speakers preferred to produce the final syllable at certain parts of the repetition cycle over others. As the speakers repeated the phrase at gradually increasing repetition rates, they fell in and out of stable, quasi-discrete rhythmic modes, and this was found for all three languages.

Another crucial observation in Figure 2 is that the most prominent distributional mode occurs near external phase 0.5, or 1/2. The speakers strongly preferred to produce the phrase-final syllable “may” halfway between the start of successive repetitions. Phases which can be expressed as simple fractions, such as 1/2, are called “simple harmonic phases”, and are frequently preferred by speakers.

The following sections focus on some of the rhythmic modes found in the three languages. Rhythmic modes were identified by taking the external phase of the phrase-final syllable of each phrase, and making a histogram of observed phases from all repetitions and speakers. (Histograms for each phrase are not included due to space limitations.) Prominent local distributional modes in each histogram were treated as rhythmic modes, i.e., stable temporal patterns for repeating the phrase. Each rhythmic mode contained all data points that were in histogram bins that were at least half as tall as the tallest bin of the mode and contiguous with it. Individual repetitions that belonged to the rhythmic mode according to this criterion were then pooled to derive the mean phase of each syllable for the particular rhythmic mode.

Figure 1: External and internal phases.
3.2. External phase

Figure 3(a)-(c) shows representative rhythmic modes found for the phrases in the three languages. The vertical lines are the beats of individual syllables; each beat is the mean external phase across all repetitions that belong to the rhythmic mode. The thick vertical lines are the beats of the phrase-final syllables.

Figure 3(a) shows rhythmic modes for the Japanese phrases. These modes share a common characteristic, in which the phrase-final syllable beats (note the thick vertical lines) occur close to external phase 0.5. Thus, for phrases of various lengths, Japanese speakers find it comfortable to repeat each phrase such that the last syllable is produced at a simple harmonic phase, such as 1/2.

The Arabic data, shown in Figure 3(b), reveals a qualitatively different result. Contrary to Japanese, the phrase-final syllable does not show a uniform tendency to occur near external phase 0.5. The final syllable for phrases 2, 4, and 6 are instead substantially later in phase than for phrases 1, 3, and 5. However, when we focus on the final stressed syllable of each phrase, we find that it is close to external phase 0.5 for all rhythmic modes shown. That is, Arabic speakers prefer to produce the final stressed syllable near a simple harmonic phase, even if it is followed by an unstressed syllable. Figure 3(c) shows analogous rhythmic modes for English. Similar to Arabic, external phase of the phrase-final syllable is later if it is unstressed than if it is stressed. However, the final stressed syllable occurs near external phase 0.5 for all phrases.

Together, Arabic and English show a similar pattern of rhythmic modes, while Japanese shows a qualitatively different result. Japanese speakers prefer to place the phrase-final syllable near a simple harmonic phase, while Arabic and English speakers do not necessarily do so. These speakers instead appear to pay closer attention to the final stressed syllable.

3.3. Internal phase

We now examine internal phase measurements in Arabic and English. Internal phase was defined relative to a “cycle” from the first to the third (final) stressed syllable of the phrase. A histogram of the internal phase of the phrase-medial stressed syllable was first made for each phrase by collapsing across repetitions measured from all metronome rates done by the speakers. Then the same criteria were used for external phase histograms to identify rhythmic modes, i.e., stable ways in which stressed syllables of a phrase are temporally patterned. Generally, the histograms showed a single rhythm mode for each phrase. The following figures are based on these rhythmic modes.

Figure 4 compares the internal phase of 4-syllable phrases in Arabic and English. Each vertical line is the mean internal phase of the syllable based on repetitions that belonged to the rhythmic mode. The two phrases showed a slight difference in the relative timing of the medial stressed syllable. The Arabic syllable “baa” is further away from the dotted line at internal phase 0.5 than is the English syllable.

Figure 3: Rhythmic modes in (a) Japanese, (b) Arabic, and (c) English.
Figure 4: Rhythmic modes in internal phase for one Arabic and one English phrase.

Figure 5: Mean internal phase of phrase-medial syllable compared within four pairs of Arabic and English phrases.

"Doug". Internal phase 0.5 is important because a syllable beat that occurs at 0.5 would be halfway between the first and third stressed syllable beats, yielding an isochronous sequence of stress beats.

Figure 5 compares the internal phase of four pairs of Arabic and English phrases that have comparable stress patterns. The figure shows means and standard deviations for just the medial stressed syllable. In the left 2 panels, the phrases have the stress pattern "ssus" or "ssusu", where the first and second stressed syllables are adjacent. In the right 2 panels, the pattern is either "ssusus" or "ssusu", where the first and second stresses are intervened by 2 unstressed syllables.

The mean internal phase deviates from 0.5 in different directions between the left and right panels. This is caused by differences in the number of unstressed syllables between the first and second stresses (0 vs. 2). In either case, the magnitude of deviation is consistently larger in Arabic than in English, for all four sets of comparisons. The differences were statistically significant. This indicates that English speakers more closely approximated stress beat isochrony within a phrase than did Arabic speakers. Even though both languages have been classified as "stress-timed", these data indicate that there are subtle yet systematic differences. It appears that English is more "stress-timed" than Arabic under these conditions.

4. CONCLUSION

Analysis using external phase demonstrated that rhythmic modes in Arabic and English are similar to each other, while different from Japanese. Arabic and English speakers seem to pay close attention to the stressed syllables, producing them at simple harmonic phases. This result is consistent with the notion that these languages are "stress-timed". By contrast, Japanese speakers tend to place the phrase-final syllable at simple harmonic phases. Despite their both being "stress-timed", Arabic and English also show systematic differences, according to internal phase measurements. Stressed syllables within a phrase deviated from a strictly isochronous sequence to a greater extent in Arabic than in English.

Overall, this study demonstrates that there are phonetically observable differences and similarities in rhythmic organization across languages. Some of these are captured by typological distinctions such as "stress-timed" vs. "syllable-timed", but there are systematic differences even among languages traditionally labeled "stress-timed".

ACKNOWLEDGEMENTS

We would like to thank Robert Port and Ken de Jong for their comments. We would also like to thank the audience at the 1998 Indiana University Linguistics Lunch Series and at the 1998 Mid-continental Workshop on Phonology.

NOTES

1. For the English phrase "great Bombay demons", internal phase of the medial syllable "Born-" showed 2 rhythmic modes, one close to internal phase 1/3, and one close to 1/2. The latter was selected for analysis because its phase better paralleled those identified for the other English phrases.

REFERENCES