Using Reiterant Speech to Study Prosodic Phenomena in Thai

Jack Gandour, Siripong Potisuk, and Judy Perkins
Purdue University

The study of prosody can be extremely difficult because of segmental variations from one syllable to the next throughout the course of a normal utterance. Some sounds are intrinsically short or long, soft or loud, low or high pitched (Lehiste, 1970). The variations that the sounds cause in timing, amplitude, and fundamental frequency ($F_0$) confound the measurement of prosodic features. The use of nonsense syllables in the study of prosodic phenomena, so-called “reiterant speech,” has been developed to circumvent these difficulties (Liberman, 1978; Liberman & Streeter, 1978; Nakatani & Shaffer, 1978). This speech is obtained by substituting *ma*, or some other nonsense syllable, for every syllable of a meaningful sentence. By using the same *ma* syllable everywhere in a sentence, prosodic regularities are subject only to the influence of factors such as stress and constituent structure. Another advantage to reiterant speech is that it is easily segmentable on spectrograms. In the case of *ma*, the nasal murmur (/m/) is different from the oral resonances (/a/) both in amplitude and spectral structure.

Reiterant speech has been used to study both acoustic and perceptual characteristics of prosody in English (Liberman & Streeter, 1978; Nakatani, O’Connor, & Aston, 1981; Larkey, 1983) and Swedish (Carlson, Granstrom, Lindblom, & Rapp, 1973). Findings indicate that reiterant speech can be a powerful and effective tool for prosody research. Some individual speakers, however, are unable to produce good reiterant speech. It is recommended therefore that potential speakers be pretested so that further experiments on the perception and production of prosody are based on reiterant speech from those speakers who are able to echo the prosody of normal sentences.

As far as we know, no one has attempted to use reiterant speech techniques for studying prosodic phenomena in a tone language. Moreover, studies so far have focused on timing patterns exclusively. But in the measurement of $F_0$, segmental and suprasegmental variations are also confounded. Perturbations of $F_0$ in vowels after voiced and voiceless consonants and intrinsic differences in $F_0$ that vary with vowel-height variations are well documented (Hombert, Ohala, & Ewan, 1979; Ohala, 1978). Larkey (1983) argues that it is not possible to examine differences in $F_0$ contours between nonreiterant and reiterant versions of the sentences “because this would require a perceptually motivated metric for comparing contours, which we do not have” (p. 1343). At present, we contend that a perceptually motivated metric is available for comparing $F_0$ contours. This metric requires that Hertz values be converted to an equivalent-rectangular-bandwidth-rate (ERB) scale, a psychoacoustic scale that gives equal prominence to excursions in different pitch registers (Hermes & van Gestell, 1991) and, in turn, to a $z$ score scale (Rose, 1987). In a tone language like Thai, each syllable carries a lexically specified tone.
Using this metric, it is possible to compare lexical tones of corresponding syllables in both the nonreiterant and reiterant versions of target sentences. Although reiterant speech eliminates segmental influences on $F_0$ and duration, it is important to determine whether it imposes an unnatural rhythmic structure that is not reflective of normal speech. Sentences with surface-structure ambiguities provide a direct test of how well reiterant speech preserves prosodic distinctions. As Larkey (1983) points out, “one can directly compare the corresponding syllables in the two interpretations without considering intrinsic duration and $F_0$ differences, because the corresponding syllables are the same syllables” (p. 1338).

Before using reiterant speech to explore prosodic patterns underlying surface-structure ambiguities, however, it is first necessary to establish that individual speakers can use nonsense syllables to mimic the prosodic structure of normal sentences. This preliminary step entails the evaluation of reiterant speech in nonambiguous sentences. Accordingly, in this paper we attempt to determine whether Thai speakers can produce reiterant speech, and whether nonambiguous sentences that exhibit the same tone pattern, stress pattern, and constituent structure yield the same duration and $F_0$ patterns in reiterant speech regardless of segmental composition.

**METHOD**

**Subjects**

Three native speakers of Thai provided the speech data for this pilot project. All three were graduate students at Purdue University: 1) SI, male, 32 years old; 2) SU, female, 32 years old; 3) SH, male, 28 years old. None of them had any previous experience with reiterant speech. All three were paid for their participation in the study.

**Materials**

Three pairs of sentences of increasing length—four, five, and seven syllables—were designed to test whether individual speakers neutralize intrinsic segmental duration and $F_0$ differences in reiterant speech (Table 1). Sentences (a) and (b) were identical in length, syntactic structure, tonal pattern, stress, rhythm, and intonation. They differed only in intrinsic segmental differences. All words in the three pairs were monosyllables. Fourteen other sentences were designed to be used in judges’ rating of reiterant speech only (Appendix). These sentences varied in length, syntax, and prosody, and included bisyllabic as well as monosyllabic words.

The target sentences in the three pairs in normal speech were constructed to maximize ease of segmentation from a spectrographic display. Word or syllable boundaries were marked by either stop consonants, nasals, or fricatives.

Of the five “ma” syllables representing each of the five lexical tones, three corresponded to actual Thai words (/maa/ ‘come,’ /máa/ ‘horse,’ /måa/ ‘dog’).
The other two "ma" syllables, /màa/ and /mâa/, correspond to possible but not actually occurring Thai words.

Table 1. Nonambiguous Sentence Pairs

1) a. เสื้อตัวนี้เก่า / / stùa tua nǐ kēe /  ‘This shirt is strikingly beautiful.’
    b. กวายเครื่องของ / / klùaj khrùa nǐ pʰɔm /  ‘These bunches of bananas are small.’

2) a. มีทั้ง แก้ว และ ปู / / mì tʰâak jē le puu /  ‘There are snails, ground lizards, and crabs.’
    b. มีแก้ว ช้อน แสบาน / / mì kēw cʰɔɔn le caan /  ‘There are glasses, spoons, and plates.’

3) a. ถ้าต้องเป็นสี่ชั้นคงพอ / / tʰâa tɔɔ pen sii cʰán kʰɔŋ /  ‘If we make it a four-story building, that ought to be enough’
    b. ถ้าเปลี่ยนเป็นแปดครั้งคงเปลือง / / tʰâa plīaŋ pen pêt kʰráŋ kʰɔŋ pluəŋ /  ‘If we change to eight times, it might be expensive.’

Recording Procedure

Speakers were asked to read a target sentence typed in Thai script on a 5 x 8 in. card, and then after a suitable pause to imitate the target sentence by substituting a ma for each syllable in the original utterance. They were instructed to make the sentence sound the same as the original sentence except for the substitution. Also typed in Thai script on the card was the mama imitation, which provided a visual cue to the tonal pattern in the original utterance. In addition, they were instructed to attempt to preserve the rhythm and intonation of the target sentence, to say [ma]
instead of [mɪ] in unstressed as well as stressed syllables, and to maintain the same speaking rate for both the normal and reiterant versions.

Before the recording session began, the speakers practiced saying the target sentences and their reiterant versions until the investigators were satisfied that they could say them in a natural way. For the actual recording, speakers were asked to produce each target sentence from the three sentence pairs in its normal and reiterant versions ten times. Of each of the other target sentences, speakers produced each normal sentence and its reiterant imitation one time only. The target sentences for the sentence pairs and the other sentences were presented in random order rather than in separate blocks. A random order of presentation was intended to minimize changes in speaking rate and learning effects, thus maximizing the likelihood of speakers being able to produce natural sounding utterances (cf. Larkey, 1983; Liberman & Streeter, 1978). For each speaker, the total corpus contained 74 utterances (3 pairs × 2 members × 10 repetitions + 14 other nonambiguous sentences). Because of monitoring errors committed by the investigators, SH’s productions of the first sentence pair were eliminated from his corpus. Thus, only 54 of SH’s utterances were retained for subsequent analysis.

Recordings were made in a soundproof booth using a Sony ECM-66B unidirectional microphone and a Marantz PMD-420 tape recorder. Speakers were seated and wore a custom-made headband that maintained the microphone at a distance of 20 cm from the lips. There were two recording sessions separated by one week. Ten repetitions of the (a) member of the nonambiguous sentence pairs were recorded in the first session plus seven other nonambiguous sentences; in the second session, ten repetitions of the (b) member of the nonambiguous sentence pairs were recorded plus seven other nonambiguous sentences. Each session for Section I lasted about 30 minutes.

Rating Procedure

Four Thai adults served as judges to rate recordings of the 14 other sentences produced by all three speakers. These judges were told that the study concerned prosody and that the mama imitations were supposed to preserve the rhythm, tonal, and intonation patterns of the original sentence. They were asked to listen to each sentence in its normal and reiterant versions in that order, and then rate how normal the original sentence sounded, and how well the reiterant version imitated it. They were instructed not to be too analytical about acoustic properties of sentences, but rather to simply judge whether the reiterant sentences “echoed” the original sentence (Larkey, 1983, p. 1339). They rated both the normal and reiterant sentences on a scale of 1 to 5. For the normal sentences, they were to use 1 for those sounding normal, 2 for those sounding almost normal, 3 for those sounding somewhat normal, 4 for those sounding almost abnormal, and 5 for those sounding abnormal. For the reiterant sentences, they were to use 1 for those whose imitation of the normal sentence was successful, 5 for those whose imitation did not sound like the normal sentence, and 2-4 for those whose imitation approximated the original sentence to varying degrees.