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,' /mǎa/ '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. เสื้อตัวนี้เก่
/ sɯ̂a tua níi kɛ̌ɛ /
'This shirt is strikingly beautiful.'

b. กล้วยเครือนี้ผอม
/ klûaj kʰrɯa níi pʰɔ̌ɔm /
'These bunches of bananas are small.'

2) a. มีทาก แย้ และปู
/ mii tʰâak jɛ́ɛ lɛ́ puu /
'There are snails, ground lizards, and crabs.'

b. มีแก้ว ช้อน และจาน
/ mii kɛ̂ɛw cʰɔ́ɔn lɛ́ caan /
'There are glasses, spoons, and plates.'

3) a. ถ้าต่อเป็นสี่ชั้นคงพอ
/ tʰâa tɔ̀ɔ pen sìi cʰán kʰoŋ /
'If we make it a four-story building, that ought to be enough'

b. ถ้าเปลี่ยนเป็นแปดครั้งคงเปลือง
/ tʰâa plìan pen pɛ̀ɛt kʰráŋ kʰoŋ plɯaŋ /
'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 × 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 [mɑ]

instead of [mɐ] 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.

Measurement Procedure

The tape-recorded stimuli were low-pass filtered at 10 KHz and digitized at a sampling rate of 20 KHz by means of a 16-bit AD converter with a 5 V dynamic range using the KAY CSL (Computerized Speech Lab) Model 4300 installed on a Gateway 2000 486/33C microcomputer. Durations in the target utterances were measured using cursors positioned on two simultaneous spectrographic displays (8 KHz frequency range, 300 Hz bandwidth; 4 KHz frequency range, 150 Hz bandwidth). In normal-speech utterances, durations of "syllables" were taken from the onset of one syllable to the beginning of the next. Any silent interval between syllables was included as part of the preceding syllable. Combining silence with the preceding syllable made it possible to determine whether reiterant sentences preserved the same pattern of duration differences (Larkey, 1983, p.1342). In normal speech, silence is a necessary property of voiceless consonants, whereas in reiterant speech, there were few silent periods because *ma* is continuously voiced. Any silent periods would be attributed to nonsegmental factors exclusively. In reiterant utterances, durations of *ma* syllables were taken from the onset of the low-amplitude nasal murmur in one syllable to the onset of the low-amplitude nasal murmur in the next. Measurement precision in both normal and reiterant speech utterances was 3 ms. Mean duration and standard deviation were determined for each syllable in the sentence pairs for both normal and reiterant speech.

F_0 was computed directly from the waveform using a CSL algorithm that employs a time-domain approach to pitch analysis (modified autocorrelation with center clipping) with nonoverlapping variable frame length. For a particular speaker, frame length was determined by his pitch range to ensure that there were at least two complete cycles within a frame. The frame length was 20 to 25 ms for the two male speakers, 15 to 20 ms for the female speaker. F_0 contours of individual syllables were first equalized for duration on a percentage scale. They were then converted to an equivalent-rectangular-bandwidth-rate (ERB) scale, a psychoacoustic scale that gives equal prominence to excursions in different pitch registers (Hermes & van Gestel, 1991), and subsequently normalized to a *z*-score scale (Rose, 1987) in order to make direct comparisons across speakers. F_0 contours were normalized by pair within each speaker. To compare the point-to-point variability of F_0 contours for all tokens of each tone of content words from the sentence pairs in both normal and reiterant speech, an ensemble-averaging approach was taken (cf. Atkinson, 1976, pp. 441–442). All tokens of a tone made up an ensemble of F_0 contours for that tone by that speaker. Eleven measurement locations were selected in a syllable at 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% of total duration. The ensemble mean and standard deviation were determined at each 10% interval. The ensemble mean can be considered as the target F_0 contour for a particular tone and the ensemble standard deviation as a measure of the random perturbations that result in deviations from the target contour.

Evaluation Criteria

Three criteria were employed to determine whether reiterant speech preserves the suprasegmental features of normal speech unconfounded by segmental influences: 1) neutralization of intrinsic duration and F_0 differences, 2) mean difference in syllable durations between corresponding syllables in the two reiterant sentences (cf. Larkey, 1983), and 3) judges' ratings of reiterant speech. With respect to duration, the first criterion was that the 0.95 confidence intervals for each of the corresponding syllables in the two reiterant sentences must overlap one another in no fewer than *n - 1* syllables. For segmentally induced perturbations of F_0, the first criterion was that the confidence intervals for each of the 11 measurement locations of the corresponding syllables in the two reiterant sentences must overlap one another in no fewer than 9 locations. The second criterion, mean duration and F_0 differences between syllables in the two reiterant sentences, was another measure of the degree to which speakers neutralized intrinsic segmental differences. Mean differences in syllable duration of 20 ms or less and in F_0 contours of 10 Hz or less were taken to indicate that the speaker satisfied this criterion. The third criterion, judges' ratings of reiterant speech, was to determine how effectively the reiterant imitations echoed the original sentences. A mean rating was computed for each sentence for each speaker. Mean ratings of 2 or less on a scale of 1 to 5 were taken to indicate that the reiterant imitations successfully preserved the prosodic pattern of the original sentences.

RESULTS

Neutralizing of Intrinsic Segmental F_0 Differences

Average time-normalized F_0 contours of the three pairs of original sentences are displayed for one of the speakers (SI) in Figure 1, and their reiterant versions are shown in Figure 2. F_0 contours and confidence intervals were typical of the other two speakers as well. All three speakers satisfied the criterion for neutralization of segmentally-induced perturbations of F_0 in almost all of the reiterant utterances. Speaker SI failed to meet criterion in only 3 out of 16 syllables. Speakers SH and SU failed in 3/12 and 2/16 syllables, respectively. Because of the small variability between the two original sentences, the reduction in variability was comparatively small from the original to the reiterant sentences. For any of the three speakers, no more than 17% of measurement locations in the original sentences failed to show overlap of the 0.95 confidence intervals. Two of the speakers (SI, SU) reduced the failure rate to 10% in the reiterant versions; the third speaker (SH) held steady at about 14% in both original and reiterant versions.

Mean differences in ensemble F_0 values between corresponding syllables in the original pairs were virtually identical to those between corresponding syllables in the reiterant pairs for all three speakers: SI, 5 Hz; SH, 6 Hz; SU, 5 Hz.

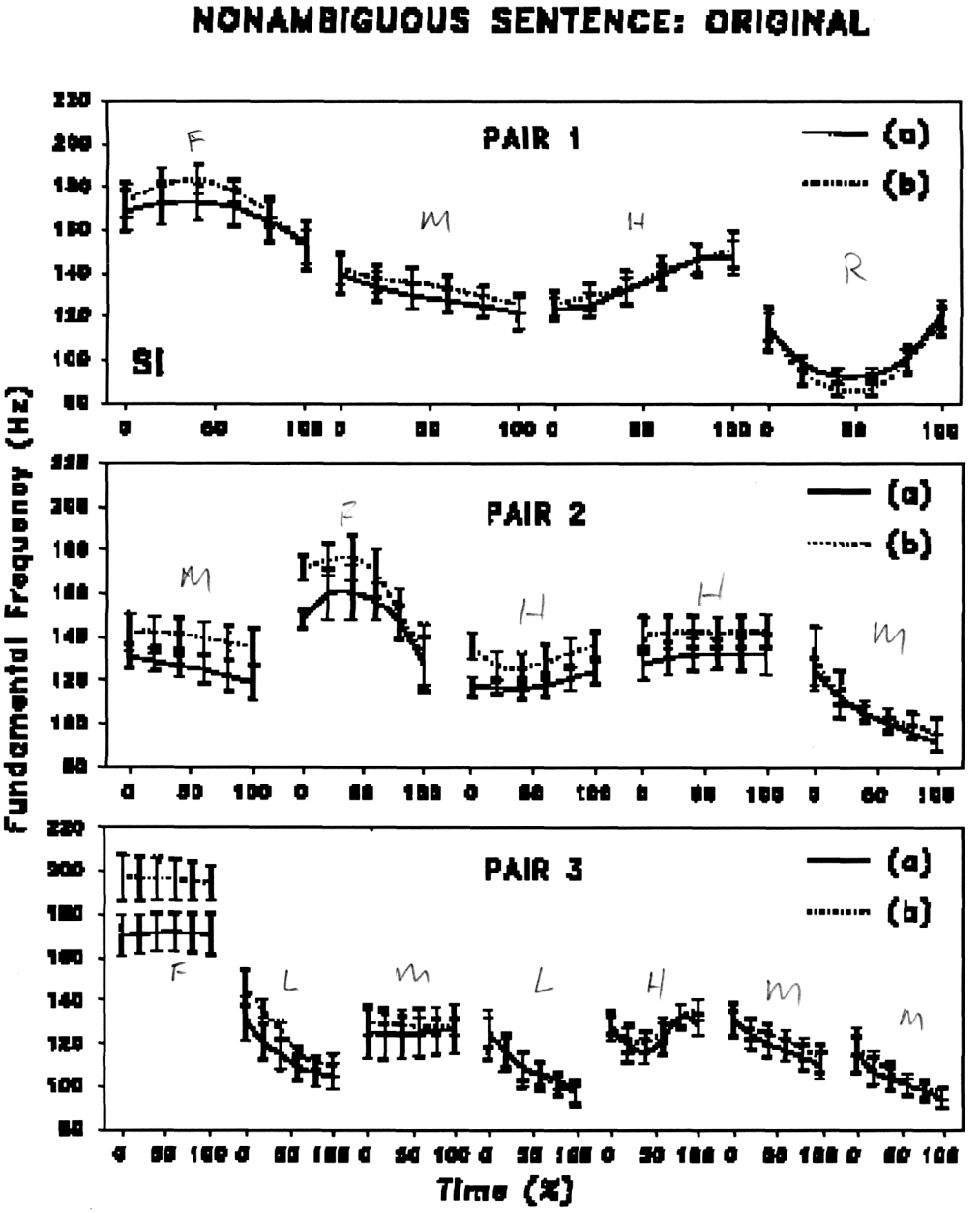

Figure 1. Original F_0 contours from one of the three speakers. A solid line represents the (a) sentence from each pair, the dotted line the (b) sentence. Error bars, which represent 95% confidence intervals, are displayed for 6 of the 11 measurement points within a syllable.

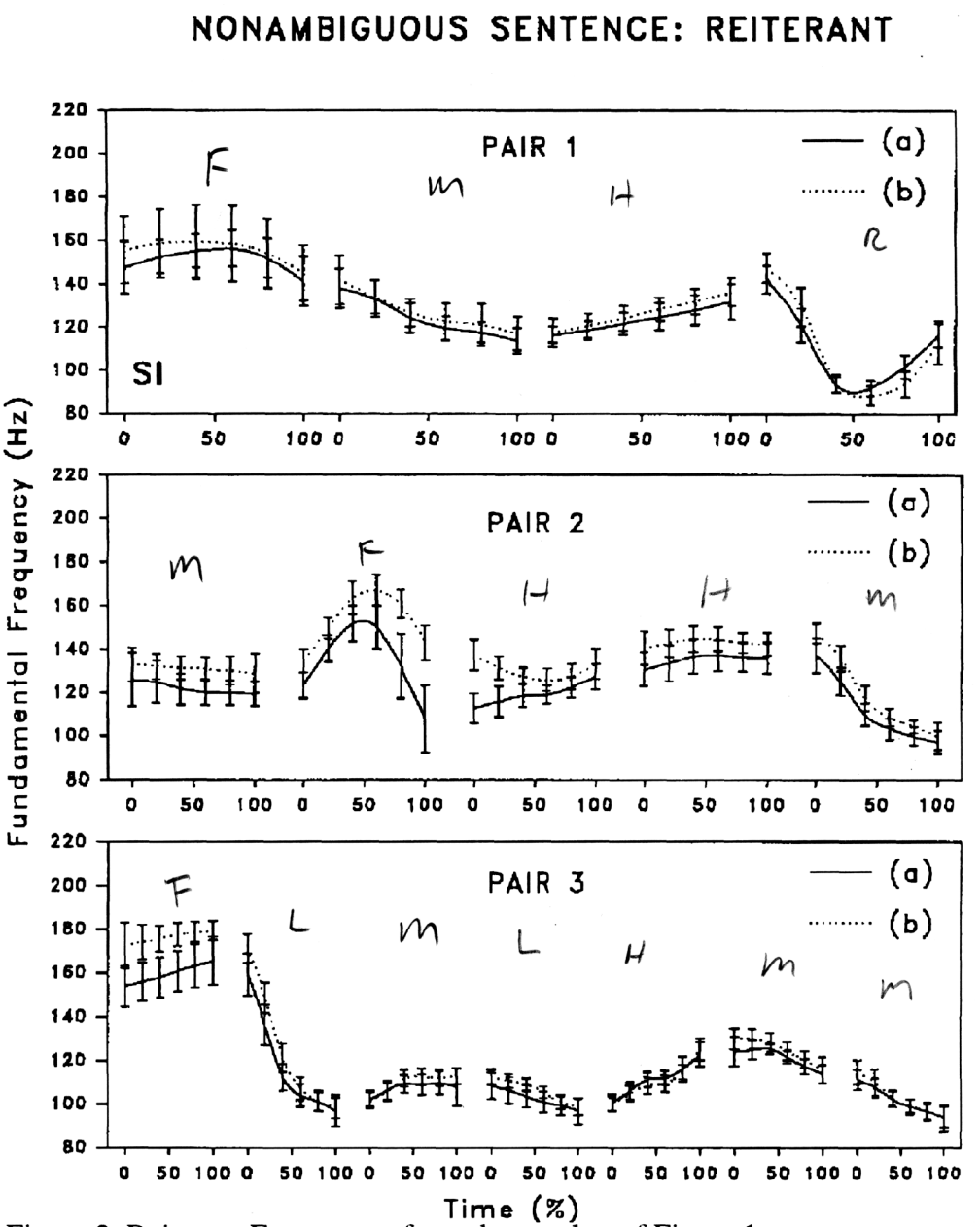

Figure 2. Reiterant F_0 contours from the speaker of Figure 1.

Neutralizing of Intrinsic Segmental Duration Differences

Syllable durations from the three pairs of original sentences are displayed for one of the speakers (SI) in Figure 3, and the reiterant versions are shown In Figure 4. SI's confidence intervals overlapped for all corresponding syllables from all

three of the reiterant pairs. SH and SU met this neutralization criterion for 1/2 and 2/3 of the reiterant pairs, respectively.

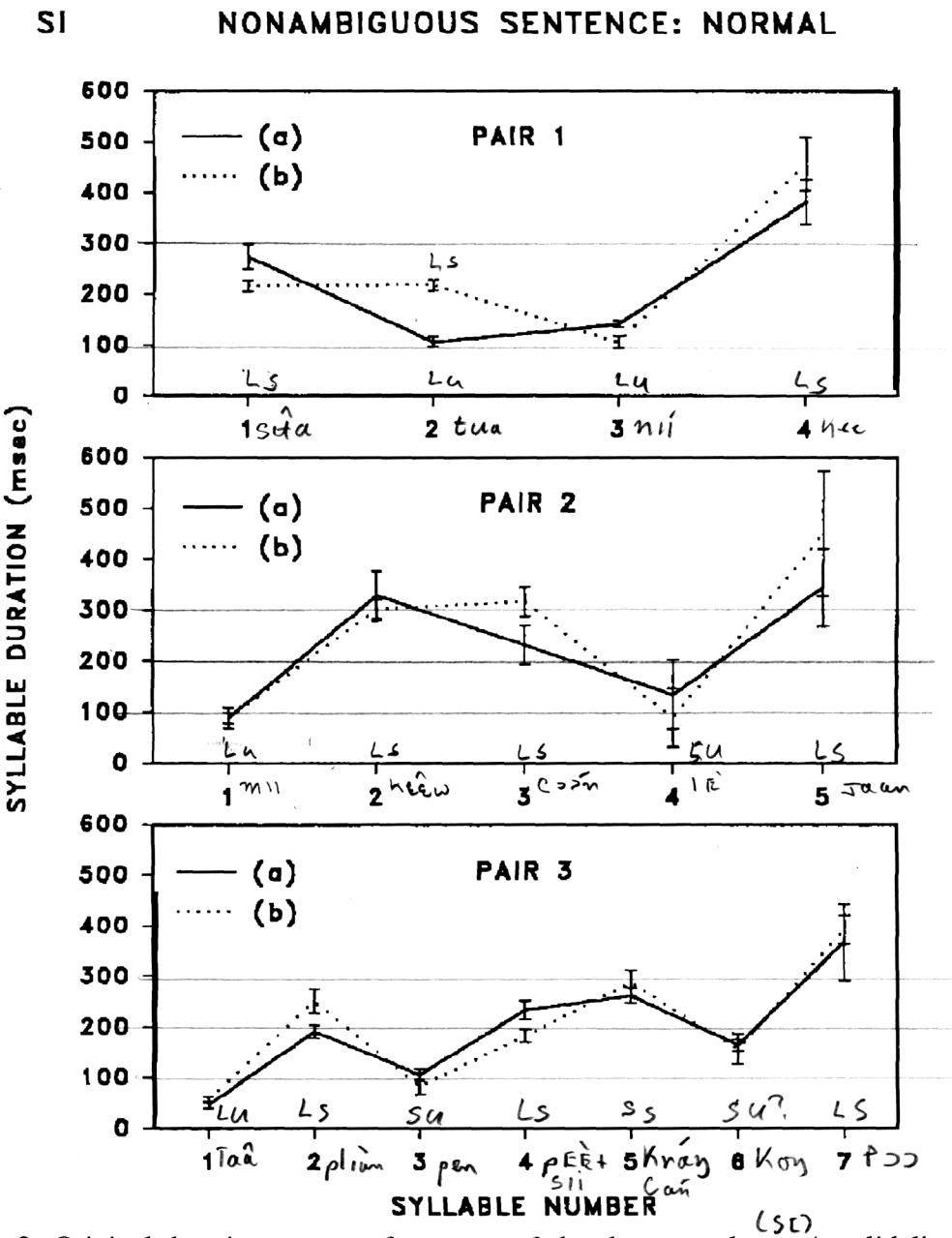

Figure 3. Original duration patterns from one of the three speakers. A solid line represents the (a) sentence from each pair, the dotted line the (b) sentence. Error bars, which represent 95% confidence intervals, are displayed for each syllable.

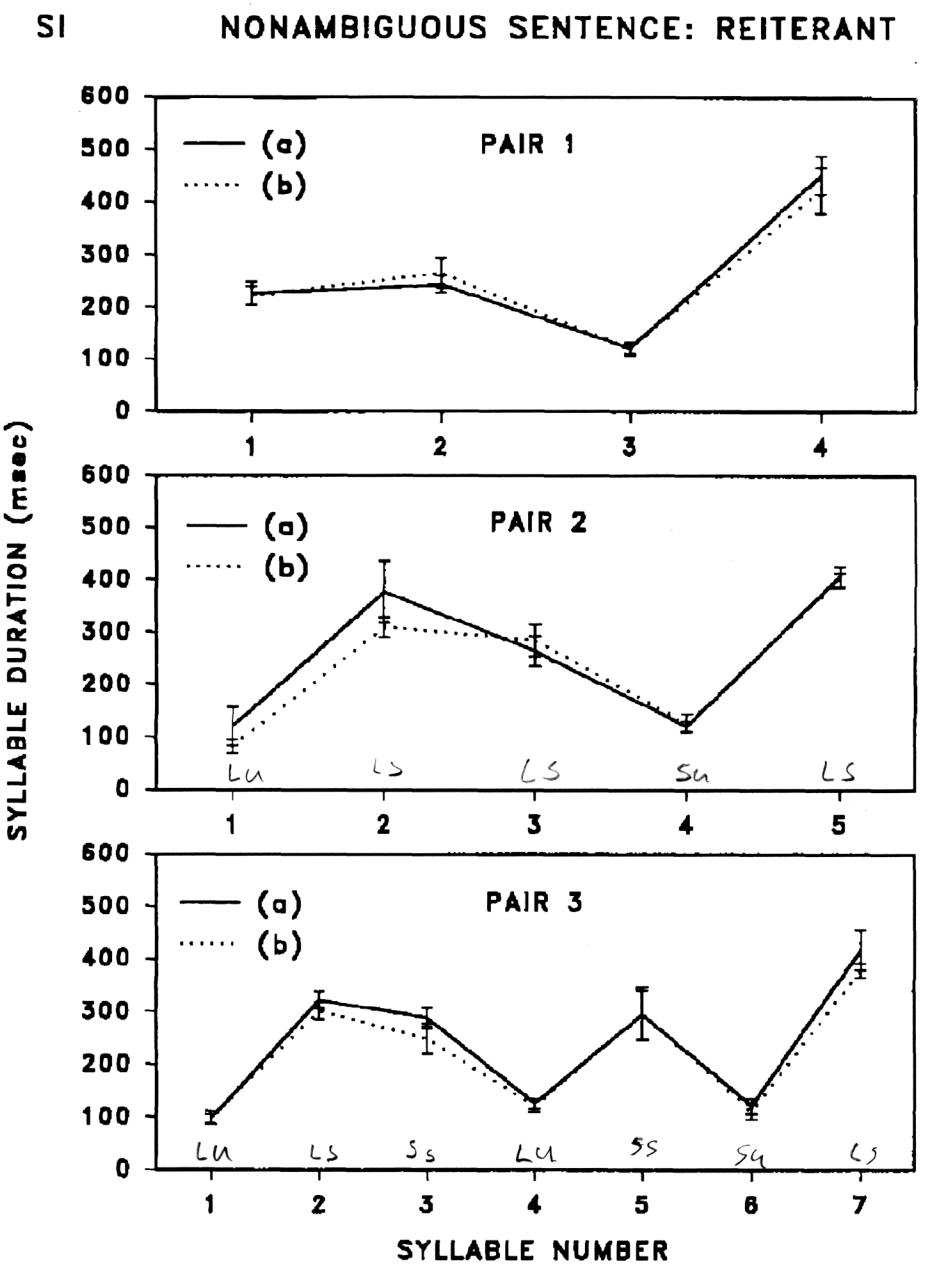

Figure 4. Reiterant duration patterns from the speaker of Figure 3. (SI)

Mean differences in syllable durations between corresponding syllables in the reiterant pairs were substantially smaller when compared to those between corresponding syllables in the original pairs for two of the three speakers: SI, $M_{original}$ = 39 ms, $M_{reiterant}$ = 19 ms; SH, $M_{original}$ = 37 ms, $M_{reiterant}$ = 19 ms. SU, however, failed to meet this neutralization criterion: SU, $M_{original}$ = 28 ms, $M_{reiterant}$ = 24 ms.

Judges' Ratings of Original and Reiterant Speech

Overall mean ratings for naturalness of the 14 additional sentences (Appendix) and accuracy of reiterant utterances were computed for each speaker (Table 2). Mean ratings of the original sentences ranged from 1.11 to 1.32, indicating that all three speakers produced natural-sounding target sentences. Mean ratings for the reiterant utterances ranged from 1.46 to 2.00, indicating that all three speakers successfully echoed the prosodic patterns of the original sentences.

Table 2. *Ratings of Original and Reiterant Speech*

	Original		Reiterant	
	M	SD	M	SD
SH	1.32	0.33	1.46	0.48
SI	1.11	0.17	2.00	0.63
SU	1.11	0.09	1.61	0.36

DISCUSSION

The findings from this preliminary study indicate that Thai speakers can use nonsense syllables to mimic the prosodic structure of normal sentences. Not all speakers, however, are equally capable of producing *mama* imitations successfully. Only one of our speakers (SI) satisfied all three criteria related to neutralizing segmental influences in reiterant speech. Both SH and SU failed to meet the overlapping criterion in terms of 0.95 confidence intervals. In the case of SH, confidence intervals failed to overlap in 3/7 syllables from the third pair; in the case of SU, confidence intervals failed to overlap in 2/4 syllables from the first pair. SU also failed to meet the criterion based on mean duration differences between corresponding syllables in the reiterant pairs. Of these three speakers, SI appears to be the best at producing reiterant speech.

Judges' overall ratings for reiterant speech are somewhat lower than those reported in Larkey (1983). The best overall score in this study was 1.46 (SH). In Larkey (1983), scores ranged from 1.08 to 1.39 across ten speakers. One factor that may account, in part, for their better performance is that her judges were linguistically sophisticated; ours were linguistically naive. Nevertheless, all three of our speakers achieved overall scores that indicate that their reiterant speech preserves prosodic patterns.

In agreement with earlier findings from other languages (e.g., Liberman & Streeter, 1978), we conclude that reiterant speech in Thai acts like nonreiterant speech in aspects which are relevant to the study of prosody. The next step in this line of research, already underway in our laboratory, is to evaluate whether reiterant speech for our two "best" speakers (SI, SH) preserves prosodic patterns in syntactically ambiguous sentences. The use of ambiguous sentences eliminates the problem of intrinsic segmental differences. In such cases, reiterant utterances are imitations of the same syllable in both readings of the ambiguous sentence. Finding

the same pattern of F_0 and duration differences in the reiterant syllables as in the normal syllables would indicate that reiterant speech not only preserves prosodic patterns, but moreover that reiterant speech can be a valuable tool in the study of prosody in Thai.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Institute on Deafness and other Communication Disorders under Grant No. DC00515-07. We wish to thank Tom Quesnell for his assistance with duration measurements.

REFERENCES

Atkinson, J. (1976). Inter- and intraspeaker variability in fundamental voice frequency. *Journal of the Acoustical Society of America, 60,* 440–445.

Carlson, R., Granstrom, B., Lindblom, B., & Rapp, K. (1973). Some timing and fundamental frequency characteristics of Swedish sentences: Data, rules, and perceptual experiments. *Quarterly Progress Status Report, Speech Transmission Laboratory, No. 4,* pp. 11–19. Stockholm: Research Institute of Technology.

Hermes, D., & van Gestel, J. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America, 90,* 97–102.

Hombert, J-M., Ohala, J., & Ewan, W. (1979). Phonetic explanations for the development of tones. *Language, 55,* 37–58.

Larkey, L. (1983). Reiterant speech: An acoustic and perceptual validation. *Journal of the Acoustical Society of America, 73,* 1337–1345.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Liberman, M. (1978). Modeling of duration patterns in reiterant speech. In D. Sankoff (Ed.), *Linguistic variation: Models and methods* (pp. 127–138). New York: Academic Press.

Liberman, M., & Streeter, L. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America, 63,* 231–233.

Nakatani, L., O'Connor, K., & Aston, C. (1981). Prosodic aspects of American English speech rhythm. *Phonetica, 38,* 84–106.

Nakatani, L., & Shaffer, J. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63,* 234–245.

Ohala, J. (1978). The production of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York: Academic Press.

Price, P., Ostendorf, S., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90,* 2956–2970.

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication, 6,* 343–351.

APPENDIX

Additional Sentences

1) ฝนเริ่มตกแล้ว
/ fŏn râəm tòk lέεw /
'It's starting to rain.'

2) แดงถูกพ่อดุ
/ dεεŋ tʰùuk pʰɔ̂ɔ dùʔ /
'Daeng is reprimanded by his father.'

3) เขาฆ่าคนตาย
/ kʰáw kʰâa kʰon taaj /
'He killed someone.'

4) อาจารย์เข้าสาย
/ ʔaacaan kʰâw săaj /
'The teacher came in late.'

5) แม่เก็บดอกไม้ขาย
/ mêε kèp dɔ̀ɔkmáaj kʰăaj /
'Mother sold the flowers she picked.'

6) น้องเกือบโดนรถชน
/ nɔ́ɔŋ kɯ̀ap doon rót cʰon /
'My brother almost got hit by the car.'

7) น้ำต้มแล้วหรือยัง
/ náam # tôm lέεw rɯ́ jaŋ /

8) ครูสอนหนังสือเด็ก
/ kʰruu sɔ̌ɔn năŋsɯ̌ɯ dèk /
'Teachers teach children.'

9) แม่ให้เงินน้องสองบาท
/ mêε hâj ŋən nɔ́ɔŋ sɔ̌ɔŋ bàat /
'Mom gave brother two baht.'

10) ผู้ร้ายต้องจับขังคุก
/ pʰûuráaj # tôŋ càp kʰăŋ kʰúk /
'Criminals, better imprison them.'

11) เขาเขียนจดหมายถึงพ่อ
/ kʰáw kʰiăn còtmăaj tʰɯ̌ŋ pʰɔ̂ɔ /
'He wrote a letter to daddy'

12) เขาเก็บมะม่วงที่สุกแล้ว
/ kʰáw kèp mamûaŋ tʰîi sùk lέεw /
'He picked the mangoes that are ripe.'

13) ถ้าไปคิดว่าอยู่เดือนนึง
/ tʰâa paj # kʰít wâa jùu dɯan nɯŋ /
'If I go, I think I'll stay for a month.'

14) เพื่อนยืมเก้าอี้ไปสองตัว
/ pʰɯ̂an jɯɯm kâwʔîi paj sɔ̌ɔŋ tua /