An acoustic and perceptual evaluation of syntactically ambiguous sentences in Thai reiterant speech

Jack GANDOUR and Siripong POTISUK
Purdue University

1. Introduction

The use of nonsense syllables in the study of prosodic phenomena, so-called "reiterant speech", has been developed to circumvent difficulties of segmental variations throughout the course of an utterance (Liberman 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. Reiterant speech can be a powerful and effective tool for prosody research (Liberman, 1978; Liberman and Streeter 1978; Nakatani and Shaffer 1978; Nakatani, O'Connor, and Aston 1981; Larkey 1983).

As far as we know, only one previous study has attempted to use reiterant speech techniques for studying prosodic phenomena in a tone language (Gandour, Potisuk, and Perkins, in press). Their findings indicate that Thai speakers can use nonsense syllables to mimic the prosodic structure of normal, nonambiguous sentences. Sentences with surface structure ambiguities, however, can provide a direct test of how well reiterant speech preserves prosodic distinctions. The aim of the present study is to compare reiterant speech to normal speech using both acoustic measurements and perceptual tests on sentences with surface structure ambiguities.

2. Duration differences in normal and reiterant ambiguous sentences

Larkey (1983) has shown for English that the same pattern of duration differences between syntactically ambiguous sentences in normal speech can also be replicated in reiterant speech. The use of ambiguous sentences eliminates the problem of intrinsic segmental duration differences. Because the syllabic composition of the alternative versions of the ambiguous sentences are identical, the reiterant syllables that are compared with each other are imitations of the same syllable. It is hypothesized for Thai that duration is the primary acoustic correlate for distinguishing alternative interpretations of some of these syntactically ambiguous sentences. The same pattern of duration differences in the reiterant syllables as in the normal syllables would indicate that reiterant speech preserves suprasegmental prosodic patterns.

MON-KHMER STUDIES 25:51-68

Method

Subjects

Two native speakers of Thai provided the speech data for this preliminary investigation. Both were graduate students at Purdue University: 1) SI, male, 32 years old; 2) SH, male, 28 years old. Both speakers were able to use nonsense syllables to mimic the prosodic structure of nonambiguous sentences (Gandour et al., in press). Neither had any previous experience with reiterant speech.

Materials

Three pairs of ambiguous sentences involving alternative surface structure representations were designed to test whether individual speakers use prosody to resolve the surface structure ambiguity (see Table 1). Within each pair, sentences (a) and (b) were identical segmentally and tonally so that we could investigate the relationship between syntax and prosody independent of any differences contributed by the segments or tones. They differed only in their surface structure representations. The target words in pairs 1 to 3 are, in order, /lúuk náam/, /jàak jùu/, and /dèk k^{h}âaŋ bâan/. In version (a) of pair 1, the target words, shown in bold, include a vocative followed by a nominal subject; in version (b), the target words constitute a nominal compound. In version (a) of pair 2, the target words include a main verb followed by an auxiliary verb; in version (b), an auxiliary verb followed by a main verb. In version (a) of pair 3, the target words include a nominal object followed by a prepositional phrase; in version (b), a nominal compound. To assess the effect of utterance length on reiterant speech, sentences in the first, second, and third pairs comprised four, five, and seven syllables, respectively. Target sentences in the three pairs in normal speech were also constructed to maximize ease of segmentation from a spectrographic display. However, the third pair had to be analyzed as containing six instead of seven syllables. The last two syllables, /k^{h}on nán/, were difficult to separate reliably because of abutting, identical nasal consonants.

TABLE 1. AMBIGUOUS SENTENCE PAIRS

1) a. เสียงน้ำไหลนี่ ลูก น้ำเต็มตุ่ม
/lûuk **náam** tem tùm/

'That's the sound of water running. Son, the urn is overflowing with water.'

b. ช่วงที่ไม่อยู่ลืมปิดฝาตุ่ม เลยได้เรื่องเลย ลูกน้ำเต็มตุ่ม
/lûuk**náam** tem tùm/

'While we were gone, we forgot to cover the urn. Boy, are we in trouble. The urn is full of mosquito larvae.'

2) a. ช่วงที่นิดตั้งครรภ์ เธออยากทานแต่ของเปรี้ยว ๆ แต่ว่าแปลกนะ หลังจากคลอดแล้ว นิดอยากอยู่เหมือนเดิม
/nít jàak jùu mŭan dəəm/

'During pregnancy, Nid often had cravings for tangy-flavored things. But it's odd that after she has already given birth, she still has the same appetite.'

b. ชีวิตความเป็นอยู่ในกรุงของนิดมีแต่ความสับสนวุ่นวาย ไม่เหมือนกับตอนที่อยู่บ้านนอกเลย คิดแล้วก็น่าสงสาร นิดอยากอยู่เหมือนเดิม
/nít jàak jùu mŭan dəəm/

'Life in the city for Nid is full of havoc and confusion unlike what she used to in the countryside. What a pity. Nid wishes she could live the way she used to.'

3) a. คุณอรุณถูกจับในข้อหาล่อลวงเด็กมาฆ่า รู้สึกว่าสถานที่เกิดเหตุจะเป็นข้าง ๆ บ้านชายคนที่พบศพ สุดท้ายคุณอรุณก็รับสารภาพ เขาฆ่าเด็กข้างบ้านคนนั้น
/kʰăw kʰâa dèk kʰâaŋ bâan kʰon nán/

'Arun was arrested for the murder of a boy. I believe the scene of the crime was beside the house of the man who found the corpse. In the end, Arun confessed. He killed the boy beside that man's house.'

b. คุณอรุณปฏิเสธว่า มีส่วนรู้เห็นในการหายตัวไปของเจ้าเปี๊ยกเด็กข้างบ้าน แต่หลักฐานบ่งชัดมาที่เขาแต่เพียงผู้เดียว ในที่สุดคุณอรุณก็รับสารภาพ เขาฆ่าเด็กข้างบ้านคนนั้น
/kʰăw kʰâa dèkkʰâaŋbâan kʰon nán/

'Arun denied any involvment in the disappearance of little Piak, the boy in the neighborhood. However, all of the evidence tied him to the crime. In the end. Arun confessed. He killed that neighborhood boy.'

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

Recording Procedure

Speakers were asked to read a target sentence and its preceding disambiguating context 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 *muh* 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. The target sentences for the sentence pairs were presented in random order. 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 and Streeter, 1978). For each speaker, the total corpus contained 60 utterances (3 pairs x 2 members x 10 repetitions).

Recordings were made in a soundproof booth using a Sony ECM-66B unidirectional microphone and a Marantz PMD-420 taperecorder. 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 ambiguous sentence pairs were recorded in the first session; in the second session, ten repetitions of the (b) member were recorded. Each session lasted about 30 minutes.

Measurement Procedure

The tape-recorded stimuli were low-pass filtered at 8 KHz and digitized at a sampling rate of 20 KHz by means of a 16-bit A/D convertor 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, 1983:1342). 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.

Results and Discussion

As shown in Table 2, critical syllables and pauses are lengthened in the (a) member of each ambiguous sentence pair in both normal and reiterant speech. A series of t-tests revealed that the mean duration differences between critical syllables and pauses of the (a) and (b) members were all highly significant ($p < .001$) for both speakers, SH and SI. For these syntactically ambiguous sentences, duration differences appear to play a prominent role in distinguishing the two readings. Such duration differences are hypothesized to follow from differences in the level of syntactic representation. This is an example of how syntactic boundaries may coincide with the boundaries of major prosodic constituents. Finding these duration differences to be significant in reiterant speech further supports the notion that reiterant speech preserves important prosodic patterns.

TABLE 2. DURATION OF CRITICAL SYLLABLES AND PAUSES IN NORMAL ANDREITERANT AM BIGUOUS SENTENCES

Normal Speech (SH)

	Pair 1		Pair 2		Pair 3	
	a	b	a	b	a	b
Syllable	273(12)	72(16)	205(15)	106(13)	166(7)	148(13)
Pause	185(21)	86(11)	90(18)	101(11)	160(13)	93(51)
Syllable + Pause	458(31)	158(12)	295(16)	208(14)	326(12)	241(53)

Reiterant Speech (SH)

	Pair 1		Pair 2		Pair 3	
	a	b	a	b	a	b
Syllable	308(11)	147(24)	285(16)	137(10)	302(41)	243(21)
Pause	141(51)	2(4)	0(0)	0(0)	112(156)	0(0)
Syllable + Pause	446(53)	149(23)	285(16)	137(10)	414(182)	243(21)

Normal Speech (SI)

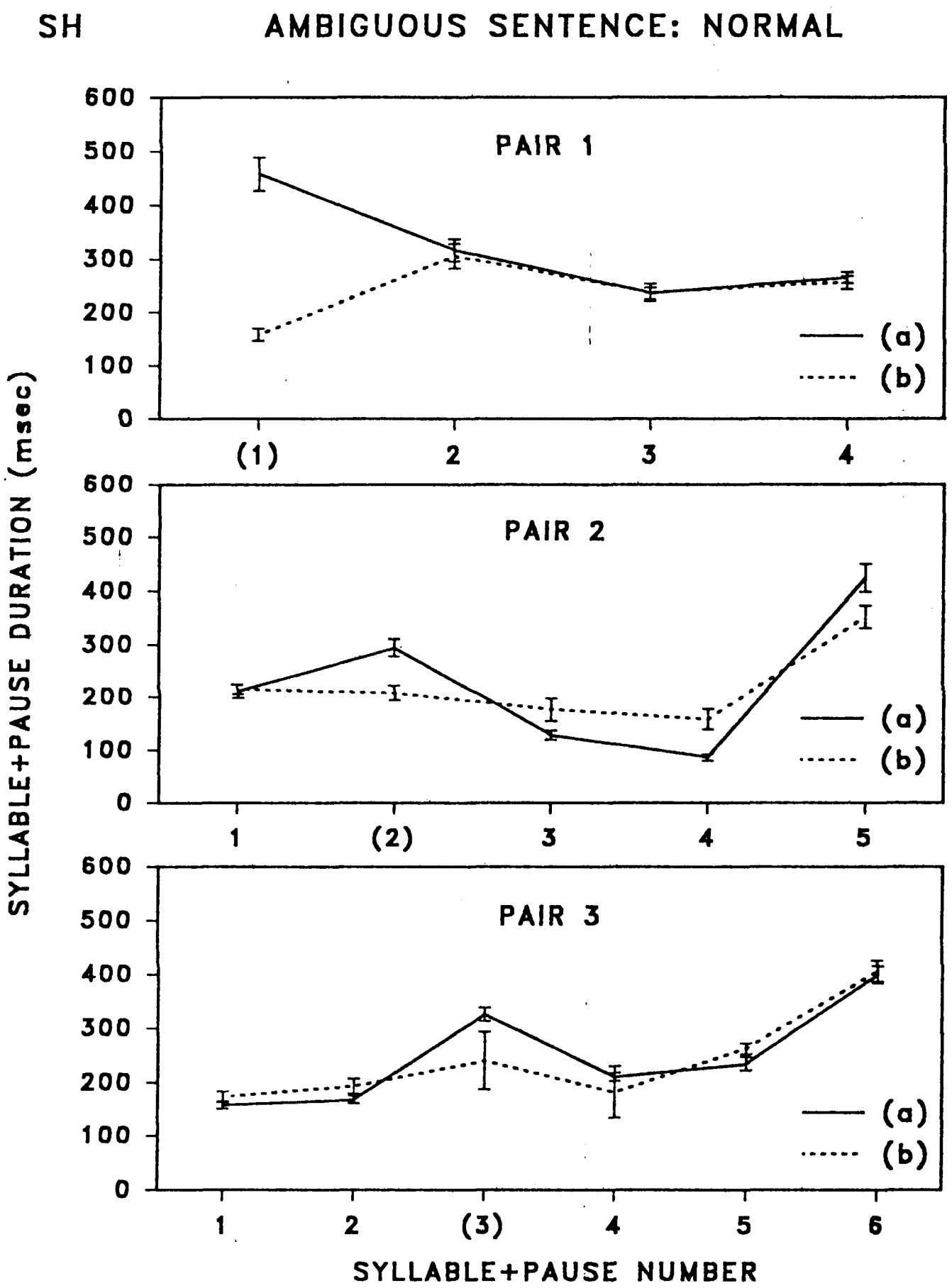
	Pair 1		Pair 2		Pair 3	
	a	b	a	b	a	b
Syllable	290(17)	174(18)	215(17)	84(10)	190(12)	140(29)
Pause	228(110)	0(0)	117(21)	88(15)	267(81)	77(14)
Syllable + Pause	518(123)	173(18)	332(28)	172(19)	457(79)	217(26)

Reiterant Speech (SI)

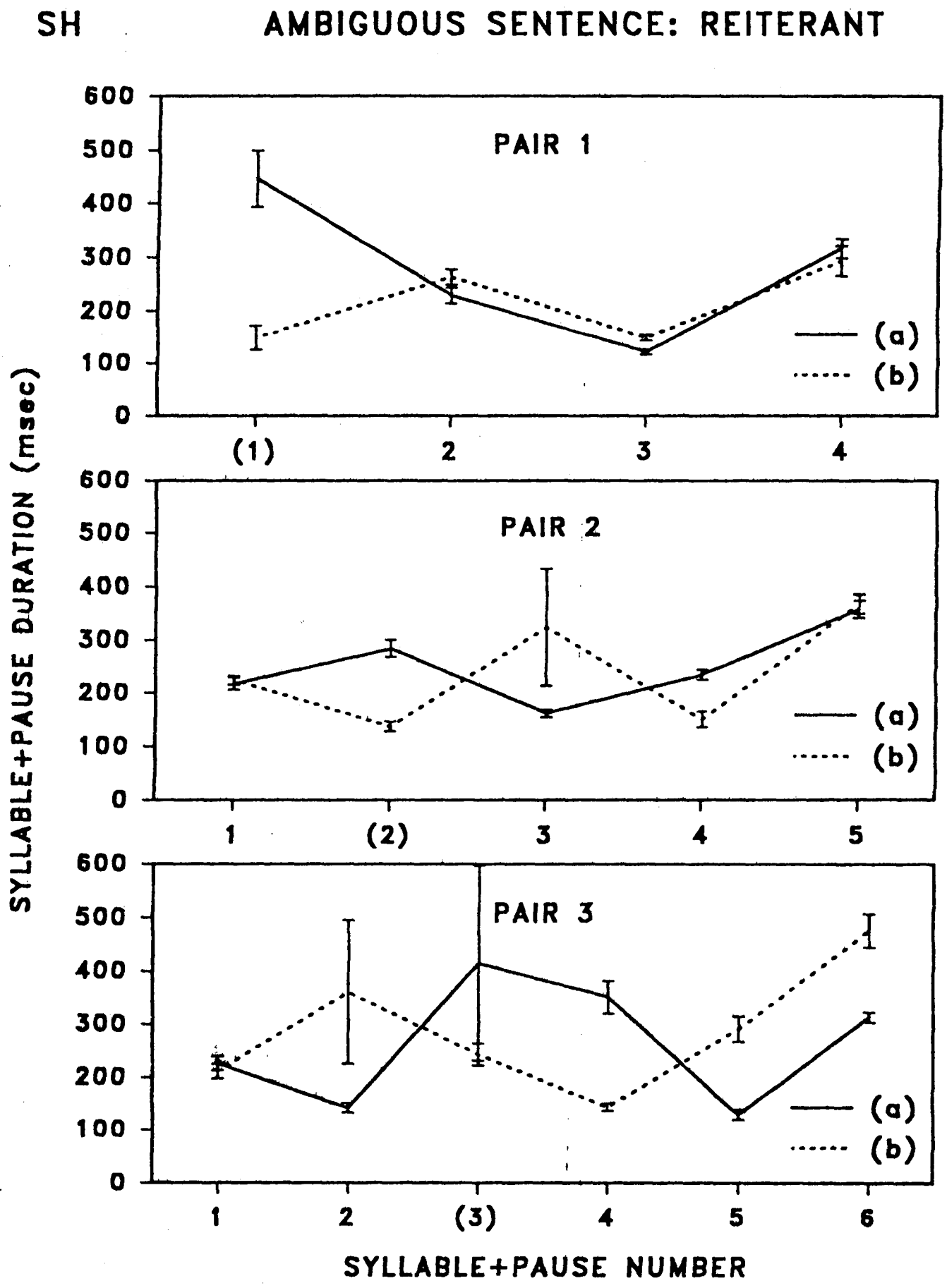
	Pair 1		Pair 2		Pair 3	
	a	b	a	b	a	b
Syllable	398(16)	121(13)	377(28)	139(14)	383(20)	255(20)
Pause	486(116)	0(0)	271(162)	0(0)	361(202)	0(0)
Syllable + Pause	884(121)	121(13)	648(174)	139(14)	744(200)	255(20)

Note. Duration values are expressed in milliseconds. Standard deviation values are enclosed in parentheses. See also Table 1.

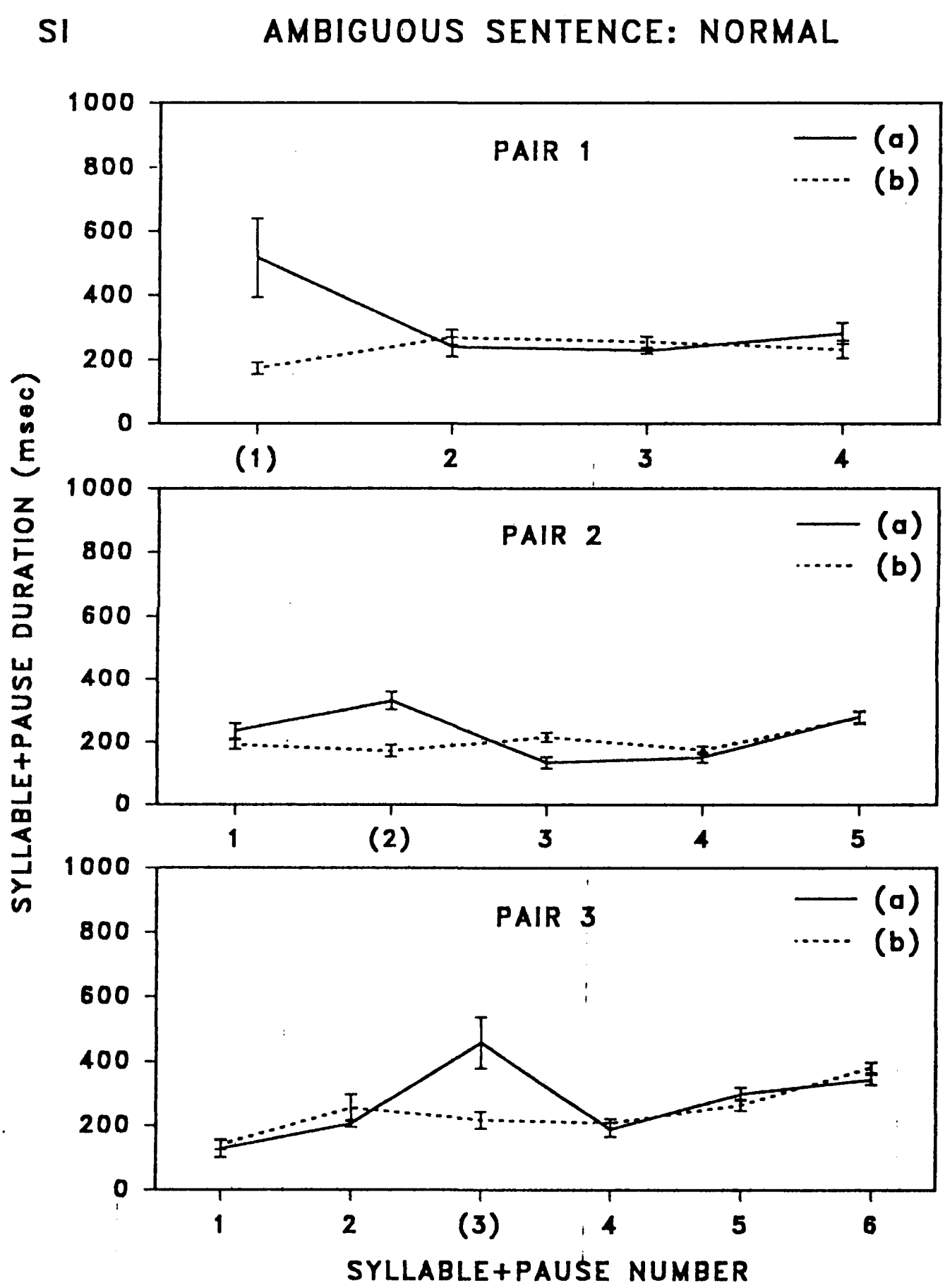
Duration patterns of the entire sentences are displayed for both speakers in Figures 1 and 2. As measured by 0.95 confidence intervals, it is observed that all corresponding critical syllables plus pause do not overlap in the two members of each normal and reiterant pair except for SH's reiterant versions of pair 3. In addition, the 0.95 confidence intervals do not overlap for /jùu/ in the normal speech versions of pair 2 as produced by either speaker. Thus, words on either side of a syntactic boundary may be varied in length to serve as prosodic cues to syntactic constituents. In reiterant speech, both speakers exhibit overlap of the 0.95 confidence intervals for syllables other than the critical one in all three pairs. Apparently, speakers employ strategies other than those used in normal speech in order to signal these syntactic distinctions in reiterant speech.



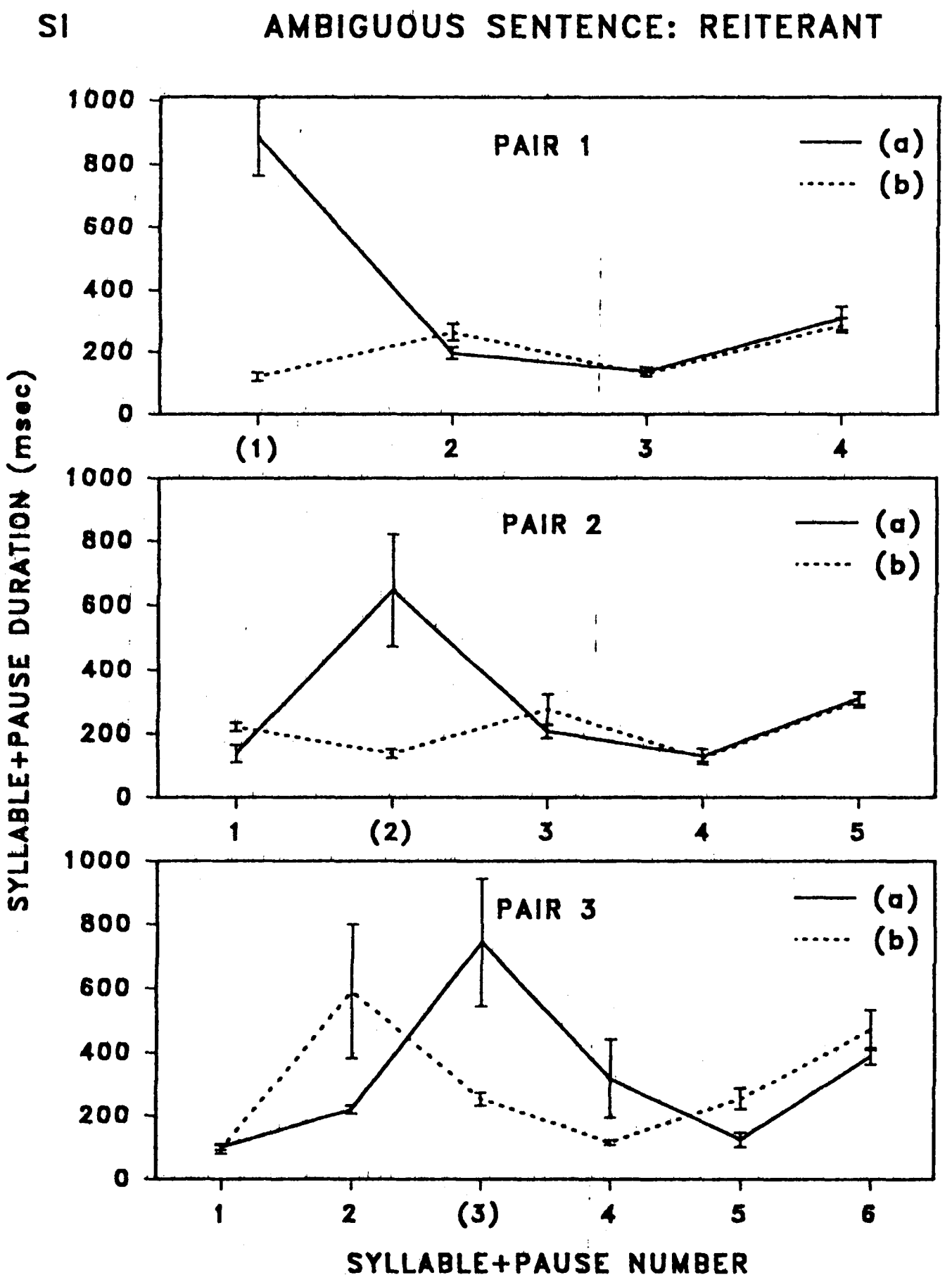
Figures 1a. Normal and reiterant duration patterns considered separately for each of three pairs of ambiguous sentences produced by speaker SH. A solid line represents (a) sentence from each pair, a dashed line (b) sentence. Error bars, which represent 95% confidence intervals are displayed for each syllable and following pause. Along the horizontal axis, the number of the critical syllable and following pause is enclosed in parentheses.



Figures 1b. Normal and reiterant duration patterns considered separately for each of three pairs of ambiguous sentences produced by speaker SH. A solid line represents (a) sentence from each pair, a dashed line (b) sentence. Error bars, which represent 95% confidence intervals, are displayed for each syllable and following pause. Along the horizontal axis, the number of the critical syllable and following pause is enclosed in parentheses.

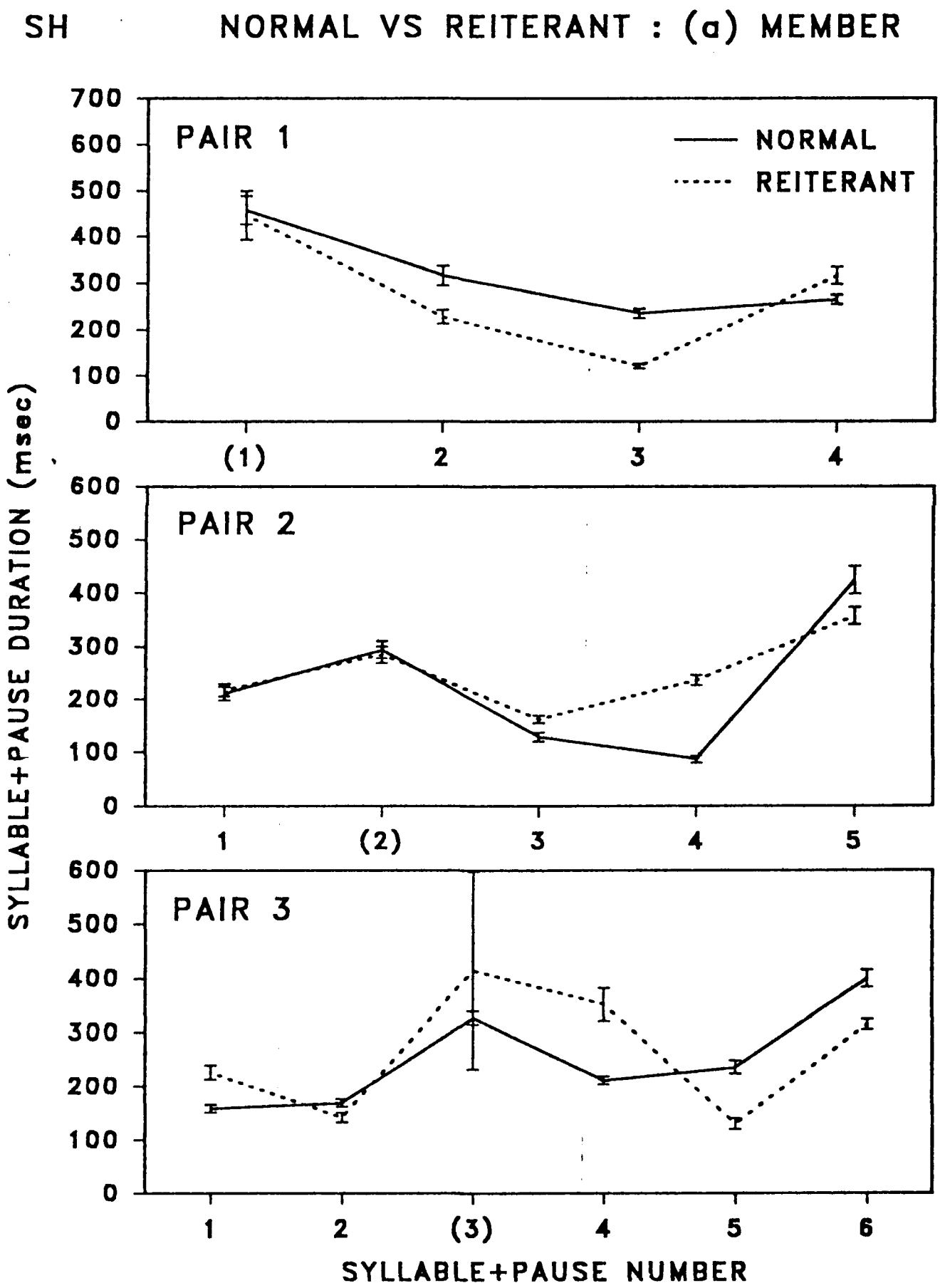


Figures 2a. Normal and reiterant duration patterns considered separately for each of the three pairs of ambiguous sentences as produced by speaker SI. See also caption to Figure 1.



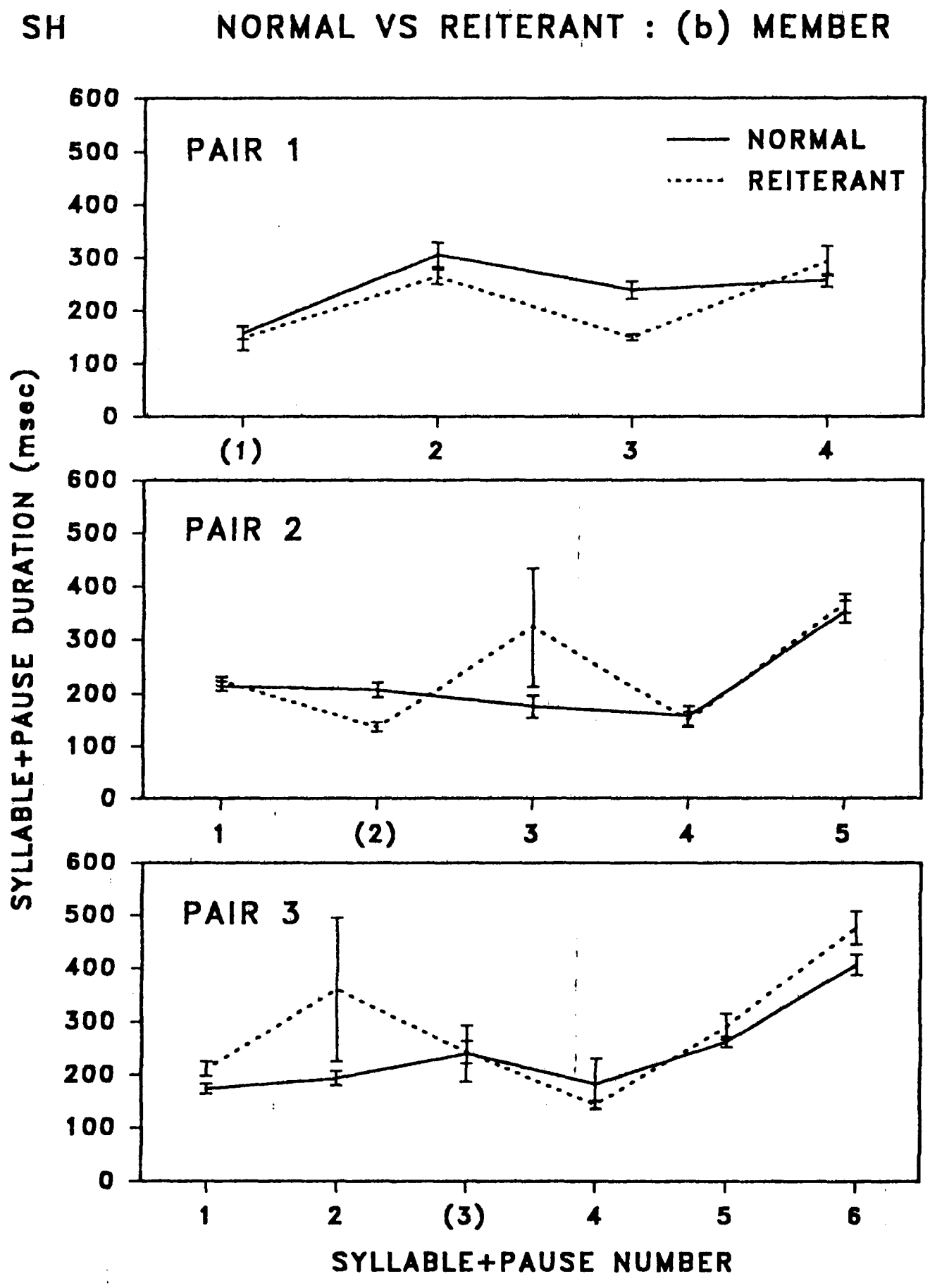
Figures 2b. Normal and reiterant duration patterns considered separately for each of the three pairs of ambiguous sentences as produced by speaker SI. See also caption to Figure 1.

Comparisons of the normal and reiterant versions of each sentence from each of the three pairs are shown for both speakers in Figures 3 and 4. In the case of SH, 0.95 confidence intervals overlap for all critical words plus pauses except for the (b) member of pair 2. SI, on the other hand, displays overlap for critical words plus pauses in the (b) member of pairs 2 and 3 only.

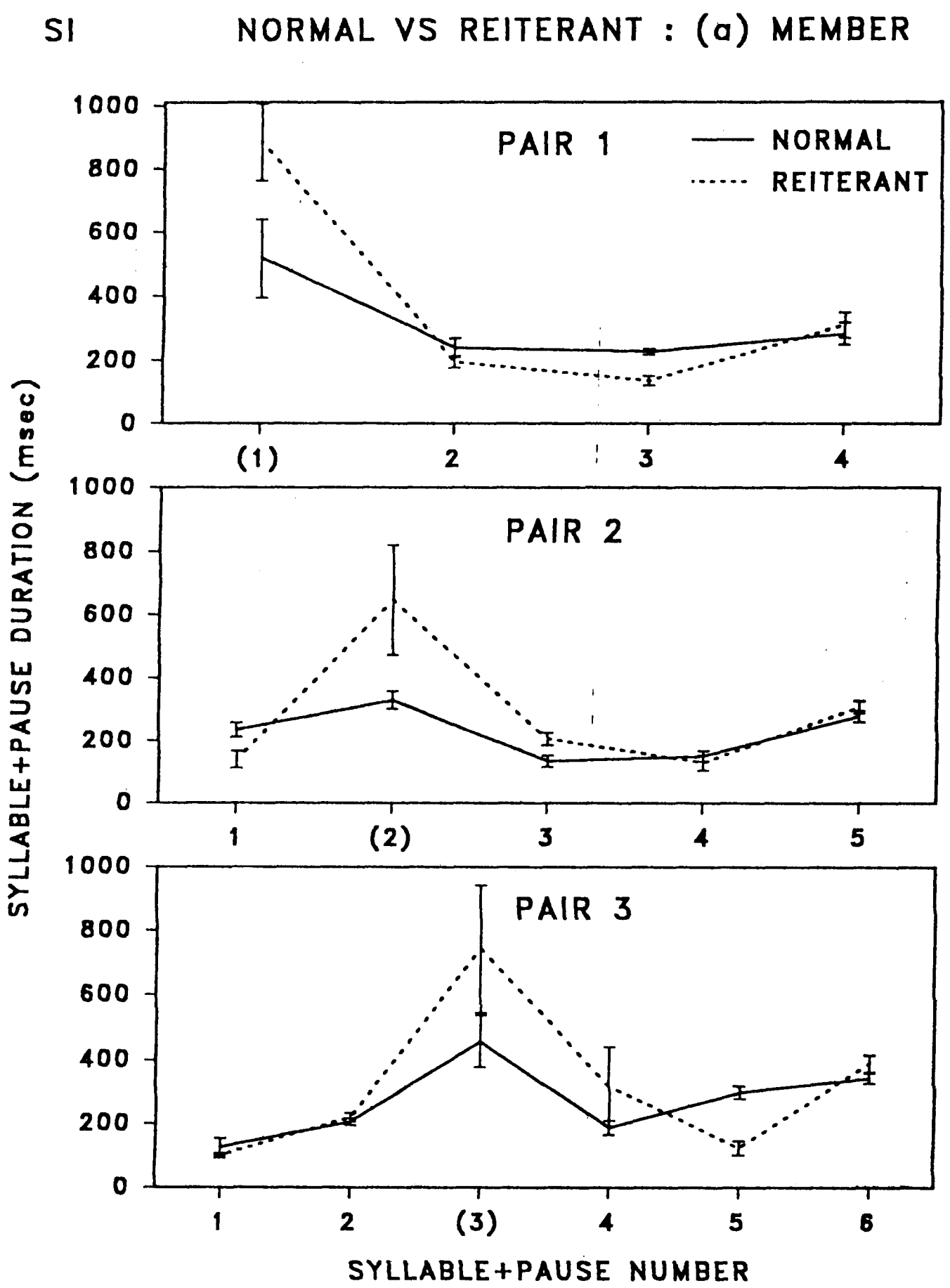


Figures 3a. A comparison of normal and reiterant duration patterns for the (a) the (b) members of the three pairs of ambiguous sentences as produced by speaker SH. A solid line represents the normal version of each member, a dashed line the reiterant version. Error bars, which represent 95% confidence intervals, are displayed for each syllable and following pause. Along the horizontal axis, the number of the critical syllable and following pause is enclosed in parentheses.

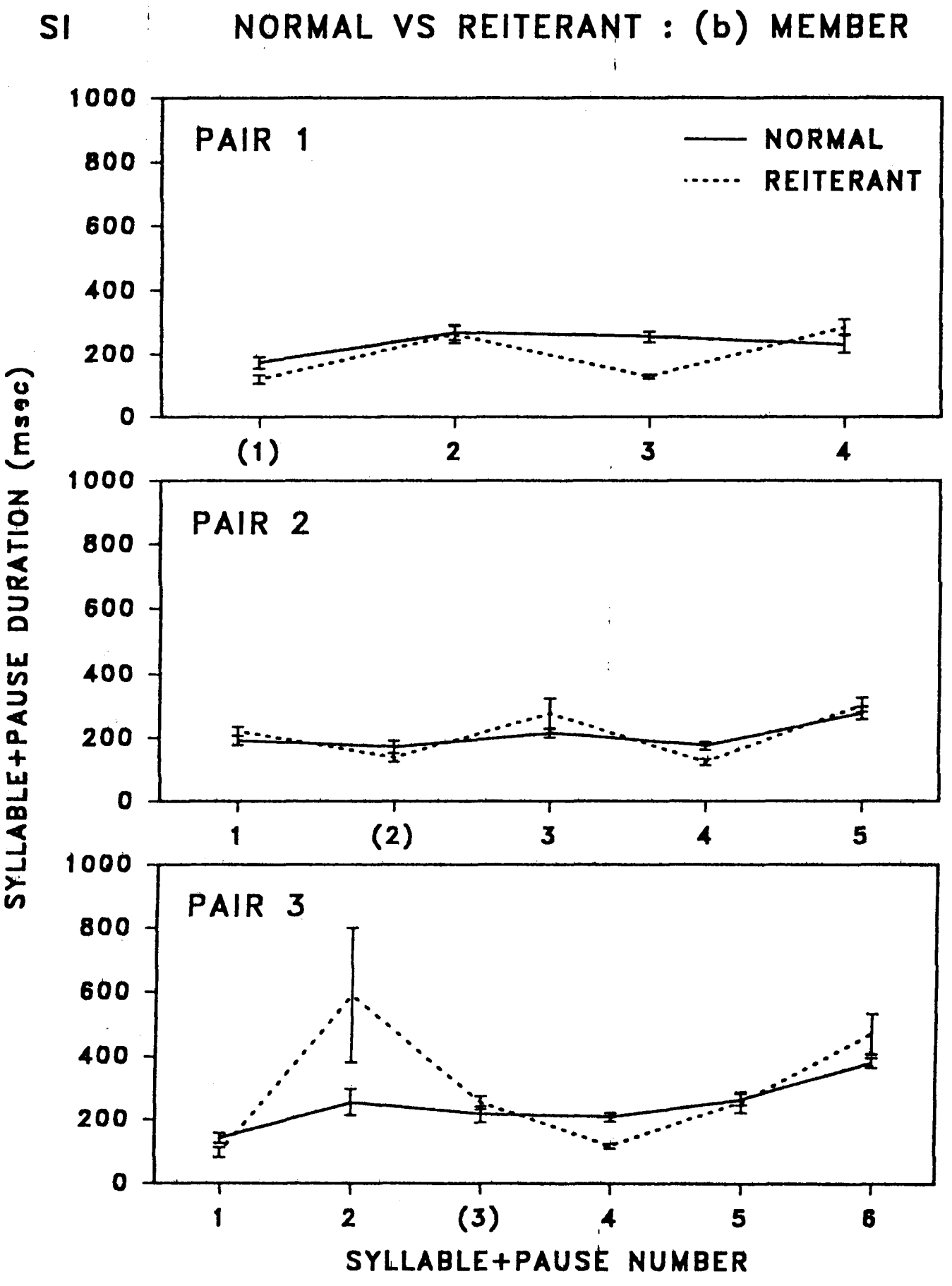
Thus, reiterant speech is at best a close approximation, but not a perfect replication of normal speech. It is possible, however, to achieve a near-perfect replication under some circumstances as exemplified by the (b) member of pair 2 as produced by SI.



Figures 3b. A comparison of normal and reiterant duration patterns for the (a) the (b) members of the three pairs of ambiguous sentences as produced by speaker SH. A solid line represents the normal version of each member, a dashed line the reiterant version. Error bars, which represent 95% confidence intervals, are displayed for each syllable and following pause. Along the horizontal axis, the number of the critical syllable and following pause is enclosed in parentheses.



Figures 4a. A comparison of normal and reiterant duration patterns for the (a) and (b) members of the three pairs of ambiguous sentences as produced by speaker SI. See also caption to Figure 3.



Figures 4b. A comparison of normal and reiterant duration patterns for the (a) and (b) members of the three pairs of ambiguous sentences as produced by speaker SI. See also caption to Figure 3.

Table 3 shows a summary of the relative amounts of lengthening found in the normal and reiterant syllables plus pauses. The results of t-tests show that relative differences in duration plays a crucial role in syntactic disambiguity in normal and reiterants speech alike. With the exception of SH's normal speech version of pair 1, the relative amounts of lengthening were substantially larger in reiterant speech as compared to normal speech. Both speakers exaggerated the durational contrasts in reiterant speech. Thus, reiterant speech clearly preserves one pattern of duration differences which has been hypothesized to be important in distinguishing Thai syntactic structures. Larkey (1983) similarly found that duration differences were preserved in English reiterant speech. However, in contrast to our findings, the relative amounts of lengthening were similar between normal and reiterant speech. Apparently, the degree to which reiterant speech "acts like" normal speech may vary depending on speaker, language, and type of syntactic construction.

TABLE 3. RELATIVE LENGTHENING OF CRITICAL SYLLABLES AND PAUSES IN NORMAL AND REITERANT AMBIGUOUS SENTENCES

		Pair 1		Pair 2		Pair 3	
		Normal	Reiterant	Normal	Reiterant	Normal	Reiterant
SH	%	190	200	42	107	35	70
	t =	30.6	13.0	13.0	24.8	4.96	2.95
	p <	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
SI	%	198	633	93	366	110	192
	t =	8.01	17.8	16.4	9.09	6.50	9.94
	p <	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Pearson product-moment correlations were computed between normal and reiterant durations of the critical words plus pauses in both the (a) and (b) for each pair. SH's correlations for the three pairs /lûuk/, /jàak/, and /dèk/ were 0.96, 0.93, and 0.52, respectively; SI's correlations were 0.82, 0.81, and 0.71, respectively. Both speakers were highly successful in preserving duration patterns of critical words and pauses in pairs 1 and 2, but somewhat less successful in pair 3, the longest sentence of the three. Sentence length, therefore, may be a limiting factor in selecting materials for use in reiterant speech experiments.

3. Disambiguation of Normal and Reiterant Ambiguous Sentences

As confirmed in the previous section, duration patterns in some ambiguous sentence pairs are preserved in reiterant speech. Although variations in fundamental frequency and amplitude contours are also expected to play a role in disambiguating ambiguous sentence pairs in Thai, duration differences are hypothesized to represent a large part of prosodic variations associated with sentence structure. In view of the robust duration differences between versions (a) and (b), it still remains

an empirical question whether or not the observed differences in duration can provide sufficient cues for listeners to choose the appropriate interpretation in both normal and reiterant speech. To the extent that listeners can identify the appropriate version at comparable levels of proficiency in both normal and reiterant speech, we have evidence to support the idea that prosodic patterns are preserved in reiterant speech.

Method

Subjects

Five Thai students at Purdue University served as judges of the ambiguous sentences produced by the two speakers. They were paid for their participation in the project.

Materials

Normal and reiterant versions of the 3 pairs of syntactically ambiguous sentences produced by the two speakers (SI, SH) served as stimuli for the perceptual experiment (see Section 2. above). For each speaker, two lists were recorded, one for normal speech and another for reiterant speech. Each list contained in random order 10 repetitions of each version of each of the three sentence pairs for a total of 60 test items. Each item on a test list was repeated three times within a trial, with a tone and two seconds of silence preceding the three repetitions of the sentence, four seconds of silence between repetitions, and five seconds between the third repetition and the tone signaling the next triple. In addition, a practice list was prepared using utterances from a third speaker. This recording was like the experimental list but with only two repetitions of each of the six sentences.

Listening Procedure

For each trial, listeners were instructed to make three judgments, one after each repetition of the sentence. Using Larkey's (1983) procedure, listeners had an answer sheet with both versions of each of the three sentence pairs, including preceding contexts, presented in order of (a) and (b), and then followed by the preceding contexts without the target sentences. After the first repetition, the listeners were instructed to circle "yes" or "no" next to version (a) of the sentence on their answer sheet to indicate whether the sentence sounded as if it were intended as version (a) or not. After the second repetition, the listeners circled "yes" or "no" next to version (b) of the sentence. After the third repetition, the listeners circled (a) or (b) to indicate a choice of which preceding context better fit the sentence that was heard.

For each speaker, listeners were presented, in order, the normal and reiterant practice lists, the normal test list, and the reiterant test list. The test tapes for both speakers were presented in a single session. Test tapes were played on a Marantz PMD420 taperecorder, and the signal was presented through TDH-39 headphones at a comfortable listening level.

Results and Discussion

Pooling across both versions of each sentence pair, both speakers, and all five listeners, the percentages of correct identification of normal and reiterant speech were 94.4% and 96.6%, respectively. As far as individual speakers are concerned, both SH and SI were highly successful in signaling the distinction between the alternative sentence interpretations in both normal and reiterant speech (SH: normal - 91.2%; reiterant - 94.4%; SI: normal - 97.7%; reiterant - 98.9%). The slightly higher identification rates for reiterant speech clearly reinforce the view that reiterant speech allows listeners to disambiguate reiterant ambiguous sentences as well as normal ambiguous sentences.

For each trial, only the choice between (a) and (b) was analyzed. Each sentence token spoken by each speaker received a score denoting the number of listeners who chose interpretation (a) of that sentence. A Pearson product-moment correlation was then computed to determine the relationship between scores on the normal sentences with scores on the reiterant sentences.

The correlations between scores on the normal and reiterant sentences for speakers SH and SI were 0.94 ($p < .0001$) and 0.99 ($p < .0001$) , respectively, indicating that the reiterant speech retained prosodic information allowing listeners to choose an interpretation of an ambiguous sentence. This demonstrates that listeners could extract syntactic information from reiterant speech comparable to that which they could extract from normal speech, thus supporting the use of reiterant speech in prosody research.

4. General Discussion

The duration measurements and the results of the perceptual tests comparing reiterant and nonreiterant ambiguous sentences suggest that reiterant speech can be a valuable tool in the study of prosody in Thai. The work presented herein can be extended in several ways. First, a larger number of syntactic structures must be examined to make the relationship between prosody and syntax more explicit. Second, acoustic measurements of amplitude and fundamental frequency should be included to establish a hierarchy of perceptual correlates that signal these differences in syntactic structure. Third, a larger number of speakers should be investigated to extend these findings to broader segments of the Thai population. Finally, for computer speech understanding applications, it is important to investigate the extension of these results to spontaneous speech where exaggeration of disambiguating strategies is expected to be minimal on the part of either the speaker or listener.

REFERENCES

Gandour, J., Potisuk, S., and Perkins, J. (*in press*). "Using reiterant speech to study prosodic phenomena in Thai." In *Studies in Linguistics in Honour of Vichin Panupong*, ed. A.S. Abramson. Bangkok: Chulalongkorn University Press.

Larkey, L. 1983. "Reiterant speech: An acoustic and perceptual validation." *JASA* 73:1337-1345.

Liberman, M. 1978. "Modeling of duration patterns in reiterant speech." In *Linguistic Variation: Models and Methods*, ed. D. Sankoff, pp. 127-138. New York: Academic Press.

Liberman, M. and Streeter, L. 1978. "Use of nonsense-syllable mimicry in the study of prosodic phenomena." *JASA* 63:231-233.

Nakatani, L., O'Connor, K., and Aston, C. 1981. "Prosodic aspects of American English speech rhythm." *Phonetica* 38:84-106.

Nakatani, L. and Shaffer, J. 1978. "Hearing "words" without words: prosodic cues for word perception." *JASA* 63:234-245.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Institute on Deafness and other Communication Disorders under Grant No. DC00515-07. We are grateful to Tom Quesnell for his assistance with acoustic measurements. Reprint requests should be sent to: Jack Gandour, PhD, Department of Audiology and Speech Sciences, Heavilon Hall, Purdue University, West Lafayette, IN 47907-1353 USA.

Received: 28 October 1994

Department of Audiology and Speech Sciences
Purdue University
West Lafayette, Indiana 47907-1353
USA

