PASAA
Vol.22
December 1992

Intraword Timing Relations in Thai

Jack Gandour, Purdue University
Sumalee Dechongkit, Mahidol University
Suvit Ponglorpisit, Prasart Neurological Hospital
Sue Young Kim, Purdue University

Abstract

Quantitative characteristics of the temporal structure of Thai words is not well documented. Even less information is available on how normal aging may affect the speech production abilities of older Thai adults. The present study examined temporal characteristics of 10 normal, young adults and 10 old adults who produced monosyllabic, bisyllabic, and trisyllabic words at a normal speaking rate. Acoustic analysis indicated that both young and old speakers reduced duration of nonfinal syllables as words increased in length. Timing patterns were generally comparable between groups whether based on absolute or relative scales of measurement. However, old speakers were generally more variable in their word productions than young speakers. Findings are discussed in relation to the effects of stress on intraword timing relations and age-related effects on speech production.

Though the phonological status of stress and rhythm in Thai was somewhat controversial in the past (Noss, 1972), the consensus opinion of linguists in more recent years is that stress and rhythm at the word level are predictable by rule (Hiranburana, 1971; Luangthongkum, 1977; Luksaneeyanawin, 1983). Minor disagreements notwithstanding, all scholars agree that the final syllable of a polysyllabic word is stressed, and that stress placement on nonfinal syllables depends on syllable structure, morphological composition of the word, and speaking rate.

Unfortunately, attention to the phonetics of stress and rhythm has not kept pace with phonological analysis. Based primarily on auditory impressions, Thawisomboon (1955) suggested that degrees of stress in Thai are correlated with relative duration of syllables. Similarly, Rudaravanija (1965) claimed that loudness and vowel duration are the phonetic correlates of stress. Hiranburana (1971) was the first to provide extensive instrumental findings on stress and rhythm at the word level. Changes in duration, segmental quality, and sometimes pitch contours are shown to vary depending on the degree of stress. Luangthongkum (1977) extended the phonetic analysis of stress and rhythm beyond the word level. Of relevance to the present study, timing data on polysyllabic words indicate a strong correlation between stress and syllable duration. The findings of both Hiranburana and Luangthongkum, however, are limited to utterances produced by single speakers. Up to the present, a large-scale database is not yet available on intraword timing relations in Thai. The aim of this study is to fill this information gap for compound words. As part of a larger series of studies on prosody in normal and brain-damaged speak-

ers, data are obtained from young and old normal adults. By including both young and old normal speakers, we are able to assess whether there are any age-related effects on syllable timing within words. It is important to determine whether any deficiencies in rhythmic patterns are to be attributed to the normal aging process rather than neurological lesions. In addition to contributing to the acoustic phonetic database on Thai, such information is also potentially relevant to issues concerning speech synthesis/recognition, aging effects on speech production, language acquisition and dissolution, and prosody instruction in Thai as a foreign language.

Method

Subjects

Twenty adult speakers of Thai participated in the study: 10 'young' adults and 10 'old' adults. These speakers were the same as those who participated in earlier studies of timing at the segment level in Thai (Gandour, Ponglorpisit, & Dechongkit, in press; Gandour, Ponglorpisit, Khunadorn, Dechongkit, Boongird, & Booklam, 1992). All 10 speakers in the young group were male and had completed 13 years of formal education; the average age of the young group was 26.4 (SD = 2.7). Five of the speakers in the old group were male, five were female; all had completed 4 years of formal education; the average age of the old group was 56.9 (SD = 2.3). All speakers had spent their entire lives in the Bangkok metropolitan area. Two different age groups as well as female speakers in the old group were included for purposes of comparison in a larger project on speech timing in brain-damaged adults.

Stimuli

Five triplets of words were chosen to investigate intraword timing relations (see Table 1). Each triplet consisted of a 1-, 2-, and 3-syllable word. The first syllable was identical in the 1-, 2-, and 3-syllable words, the second syllable in the 2- and 3-syllable words. The 2- and 3-syllable words were noun compounds. According to Thai rhythm rules (Luangthongkum, 1977; Luksaneeyanawin, 1983; Rudaravanija, 1965), primary stress was assigned to all five monosyllabic words, and to the last syllable of the polysyllabic compounds. In the 3-syllable words, the second syllable was reduced to a secondary stress at a normal speaking rate or was unstressed altogether at a faster speaking rate. To counter "incompressibility" (Klatt, 1976), whereby segments or syllables reach a point where they can be shortened no further, all nonfinal syllables ended in a sonorant. No "linker-syllables" (Bee, 1975), i.e. CV?, were present in any of the words. Syllable durations in Thai are influenced by lexical tone (Gandour, Dardarananda, & Holasuit, 1988). Tone sandhi was absent in these words. Thus, each syllable carried a single tone regardless of its position in a word. For ease of measurement, plosives were assigned to the beginning of thirteen of fifteen non-initial syllables, spirants to the remaining two.

Table 1 Word Triplets Used for Investigating Intraword Timing Relations

lăŋ 'back'	lăŋkhaa 'roof'	lăŋkhaabâan 'houseroof'
woŋ 'circle'	woŋkaan 'field'	woŋkaanphêɛt 'medical group'
sĭi 'color'	sĭidɛɛŋ 'red'	sĭidɛɛŋkhêm 'maroon'
din 'soil'	dinsɔ̌ɔ 'pencil'	dinsɔ̌ɔphɔɔŋ 'chalk'
náam 'water'	náamtaa 'tear'	náamtaathian 'candle dripping'

Recording Procedure

All words were printed in large Thai script on 3" x 5" cards. A total of 150 cards (15 words x 10 tokens) was presented in random order to subjects. They were instructed to read the words at a conversational speaking rate, i.e. at a rate they considered representative of their conversational speech. To avoid start and end effects, extra cards were placed at the top and bottom of the deck. To avoid list reading effects, a sufficient pause was provided between items to ensure that subjects maintained a uniform speaking rate. Subjects' utterances were recorded in a reasonably quiet room in a single session using a Sony ECM-66B microphone and a Marantz PMD-420 taperecorder.

Measurement Procedure

A total of 3,150 utterances were digitized at a sampling rate of 10 KHz on a Kay Elemetrics DSP Sona-Graph model 5500. A simultaneous display of a wide-band (300 Hz) spectrogram with a scale from 0-8 KHz, amplitude trace, and audio waveform was created for each utterance. Using cursors on this display, measures were taken of each syllable's duration as well as total duration of the word. Spectrograms were demarcated in time following conventional rules for segmentation of the speech signal (e.g., Klatt, 1976; Peterson & Lehiste, 1960). Measurement precision was 6 ms.

Broad phonetic transcriptions of each utterance were made by the first author and a phonetically-sophisticated native speaker of Thai. An utterance was considered to be "on-target" if both of their broad transcriptions matched that of the target word. Only the utterances judged to be on-target were retained for data analysis. In addition, those utterances contaminated by voice overlap or extraneous background noise were eliminated from further consideration. As a result of this preliminary screening, 1.6% and 2.8% of utterances produced by young and old speakers, respectively, were discarded from the corpus. Thus, the remaining number of on-target or "correct" productions that was submitted to further analysis totaled 3, 082.

To assess reliability, 5% of correct productions were randomly selected from the total corpus of 1-, 2-, and 3-syllable words across groups and speakers. For intrajudge reliability, one research assistant remeasured syllable and word durations, without access to the original measurements, after a period of two months. For interjudge reliability, a second research assistant independently measured the same set of words. Intrajudge differences between the two sets of measurements were found to average 1 ms, $\underline{r}$ = .996. Interjudge measurements had a mean absolute difference of 4 ms, $\underline{r}$ = .992.

Data Analysis

Word and syllable durations were computed in absolute and relative scales of measurement. Relative measures were essential from a linguistic and motoric perspective. Stress in Thai is signaled primarily by changes in relative temporal structure (Luangthongkum, 1977). Acoustic measures of relative speech timing remain constant across large changes in speaking rate (Weismer & Fennell, 1985). The utterances herein were part of a larger data base including those produced by brain-damaged adults. In the case of some brain-damaged speakers, speaking rates were anticipated to be slower than normal (Collins, Rosenbek, & Wertz, 1983). Without relative measures, it would not be possible to compare rhythmic patterns across speaking rates.

As a measure of stress effects, the absolute and relative duration of the first syllable was compared across all three words in a triplet-monosyllabic, bisyllabic, and trisyllabic. The absolute and relative duration of the second syllable was compared between the bisyllabic and trisyllabic words.

To evaluate relative durations expressed as proportions by means of analysis of variance (ANOVA), an arcsine transformation was used to stabilize variances (Winer, Brown, & Michels, 1991, pp. 356-357). To evaluate relative variability of durations by means of ANOVA, the coefficient of variation (standard deviation divided by the mean) was computed to compensate for any between-group or between-word length duration differences (cf. Allen, 1973; Kent & Forner, 1980; Smith, Sugarman, & Long, 1983).

To assess the magnitude of change in duration of 1st and 2nd syllables in words of increasing length, 1st syllable ratios were computed of monosyllabic to bisyllabic and trisyllabic durations; 2nd syllable ratios were computed of bisyllabic to trisyllabic durations.

Results

Absolute Duration

Words increased in duration as the number of syllables increased (see Figure 1). Monosyllabic words produced by old speakers were, on average, 4.4% longer than those of young speakers, bisyllabic words 8.8% longer, and trisyllabic words 7.5% longer. Results of ANOVA revealed a significant word length effect, $F(2,14) = 109.32$, $p < .0001$. Neither the group main effect nor the group x word length interaction reached significance. This means that both young and old speakers exhibited the same pattern of increase in word duration as words increased in length. Post-hoc Student-Newman-Keuls paired comparisons were all significant ($\alpha = .01$) between monosyllabic, bisyllabic, and trisyllabic words. For means and standard deviations of total duration of monosyllabic and polysyllabic words within groups, see the Appendix.

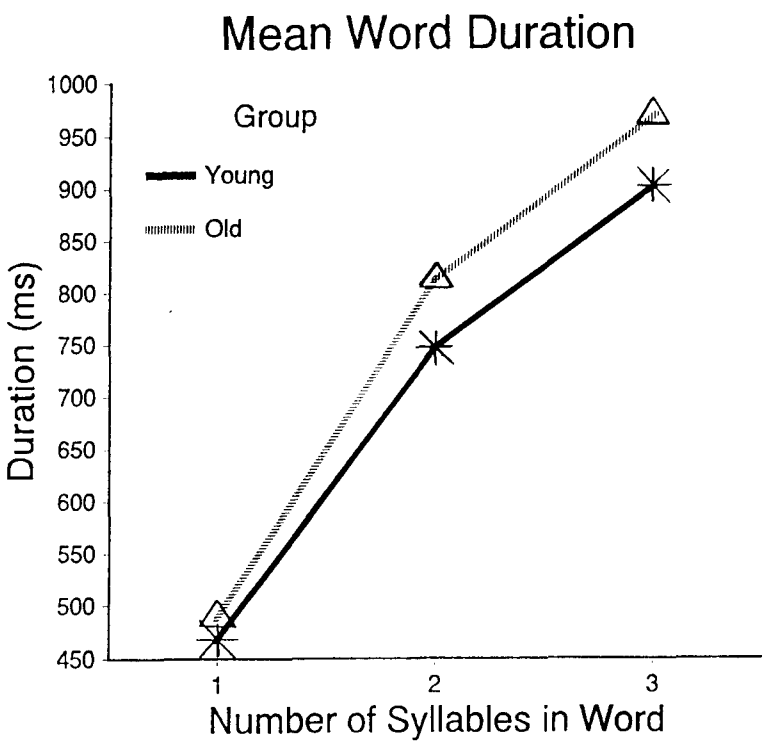

Figure 1 Mean word duration for each group as a function of word length.

The duration of the 1st syllable decreased as the number of syllables in the word increased (see Figure 2). Results of ANOVA revealed a significant word length effect, $F(2, 36) = 508.74$, $p < .0001$. Neither the group main effect nor the group x word length interaction were significant. This means that both young and old speakers exhibited the same pattern of decrease in 1st syllable duration as words increased in length. Post-hoc Student-Newman-Keuls paired comparisons were all significant ($\alpha = .01$) between 1st syllable duration in monosyllabic, bisyllabic, and trisyllabic words. Similarly, the duration of the 2nd syllable decreased as the number of syllables in the word increased (see Figure 2). The word length effect was significant, $F(1,18) = 376.06$, $p < .0001$, meaning that 2nd syllable duration in bisyllabic words was longer than in trisyllabic words. The absence of a group main effect and a group x word length interaction indicated that young and old speakers exhibited the same pattern of decrease in 2 nd syllable duration as words increased in length. For means and standard deviations of 1st and 2nd syllable durations of monosyllabic and polysyllabic words within groups, see the Appendix.

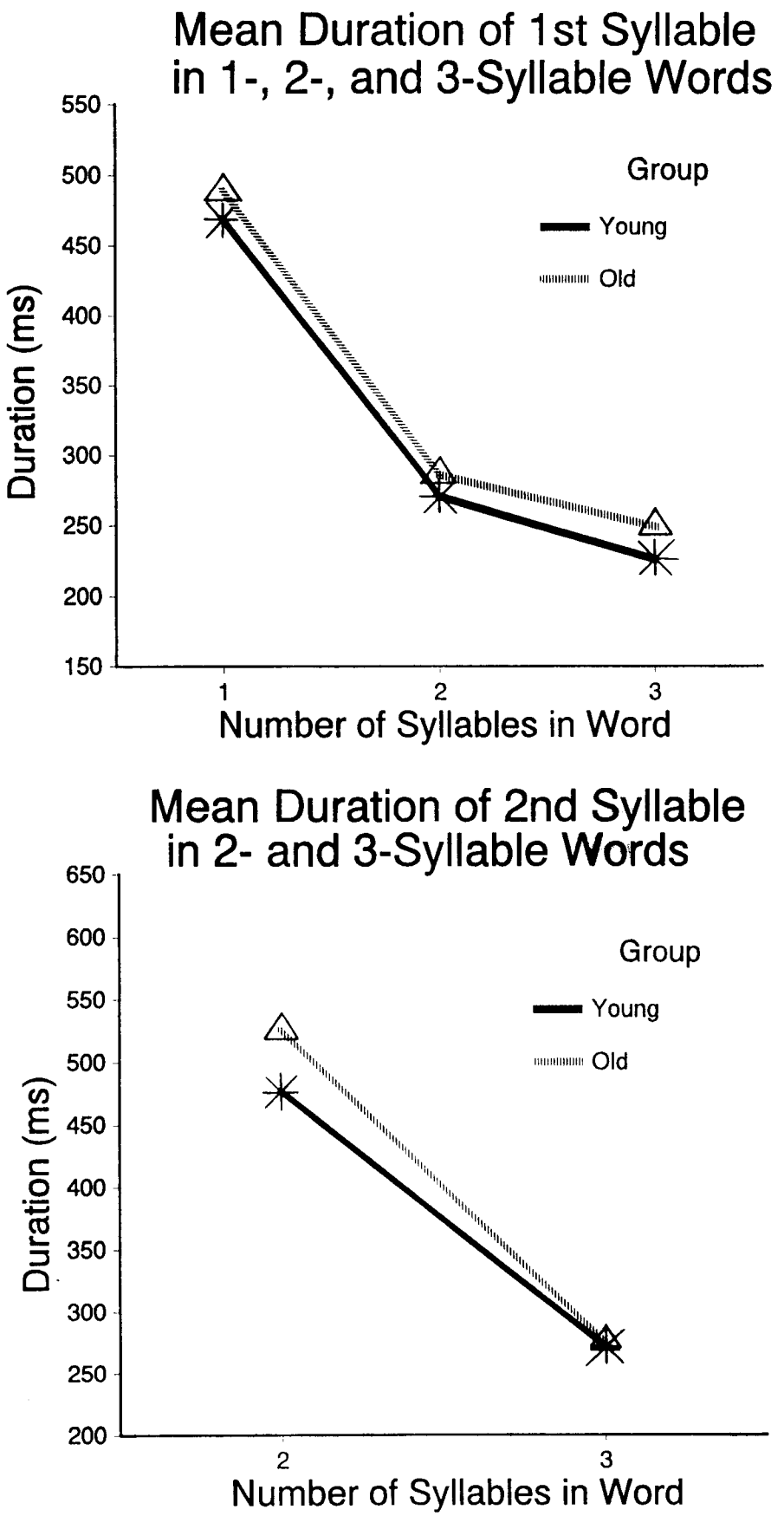

Figure 2. Mean duration of 1st and 2nd syllables for each group in monosyllabic, bisyllabic, and trisyllabic words.

Relative Duration

Expressed as a percentage of the total word duration, the durations of the 1st and 2nd syllables decreased as the number of syllables in the word increased for both groups of speakers (see Figure 3). Results of ANOVA of 1st syllable duration and 2nd syllable duration in 2- and 3-syllable words revealed significant word length effects, $F(1,18) = 990.53$, $p < .0001$ and $F(1,18) = 896.56$, $p < .0001$, respectively. In either analysis, the group main effect and group x word length interaction were not significant. This means that both groups responded in essentially the same manner to increases in word length. Pooled across groups, the 1st syllable in bisyllabic and trisyllabic words occupied 35.7% and 25.4% of word duration, respectively; the 2nd syllable in bisyllabic and trisyllabic words occupied 64.2% and 29.4% of word duration, respectively. For means and standard deviations of 1st and 2nd syllable proportions of bisyllabic and trisyllabic words within groups, see the Appendix.

Mean ratios of 1st syllable duration in monosyllabic words to 1st syllable duration in bisyllabic and trisyllabic words are displayed in Figure 4. Results of ANOVA indicated no significant differences between groups in magnitude of decrease in 1st syllable duration. Pooled across groups, 1st syllable duration decreased by 80% in bisyllabic words and 110% in trisyllabic words when compared to 1st syllable duration in monosyllabic words. From bisyllabic to trisyllabic words, there was an additional 17% decrease in the duration of the 1st syllable.

Mean ratios of 2nd syllable duration in bisyllabic to trisyllabic words are displayed in Figure 5. Results of ANOVA revealed a significant group effect, $F(1,90) = 13.28$, $p < .0004$. For young speakers, the magnitude of decrease of 2nd syllable duration was 78%; for old speakers, 97%.

Mean coefficients of variation for monosyllabic, bisyllabic, and trisyllabic words were 5.9, 4.8, and 4.8, respectively, for young speakers; 8.9, 5.6, and 5.7, respectively, for old speakers. Results of ANOVA yielded a significant two-way interaction between group and syllable condition, $F(2,294) = 9.18$, $p < .0001$. Subsequent analysis of simple main effect of word length with post-hoc Student-Newman-Keuls comparisons indicated that monosyllabic words produced by young and old speakers were significantly ($\alpha = .01$) more variable in production than either bisyllabic or trisyllabic words. Differences in variability between bisyllabic and trisyllabic words failed to reach significance for either group. Subsequent analysis of the simple main effect of each group indicated that old speakers were significantly more variable than young speakers in their production of monosyllabic, $F(1,98) = 37.71$, $p < .0001$, and trisyllabic words, $F(1,98) = 4.90$, $p < .05$, only.

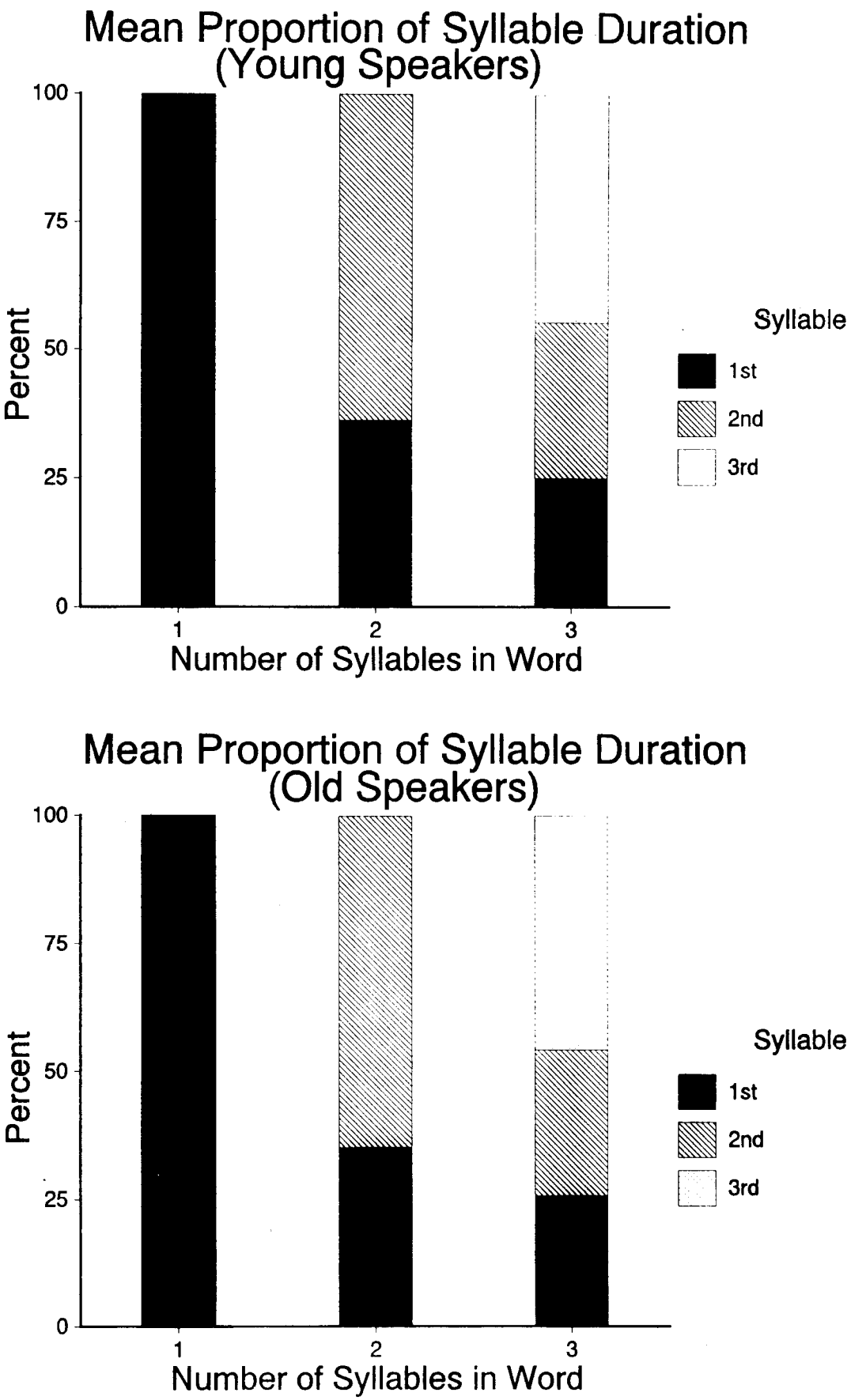

Figure 3. Mean proportion of syllable duration for each group in monosyllabic, bisyllabic, and trisyllabic words.

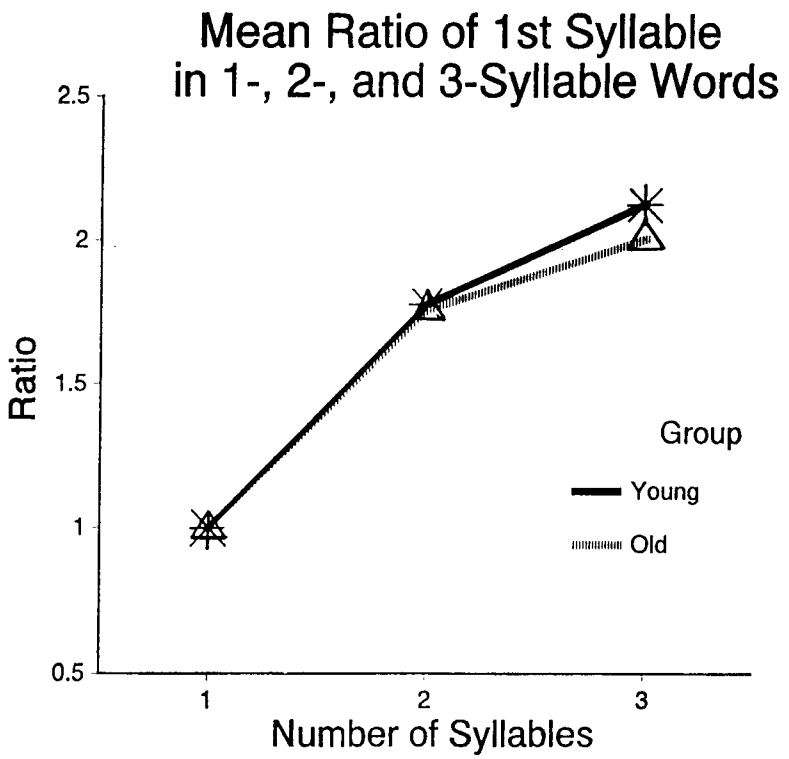

Figure 4. Mean ratio of 1st syllable duration in monosyllabic words to 1st syllable duration in polysyllabic words for each group. Ratio of monosyllabic duration to monosyllabic duration is equal to 1.0.

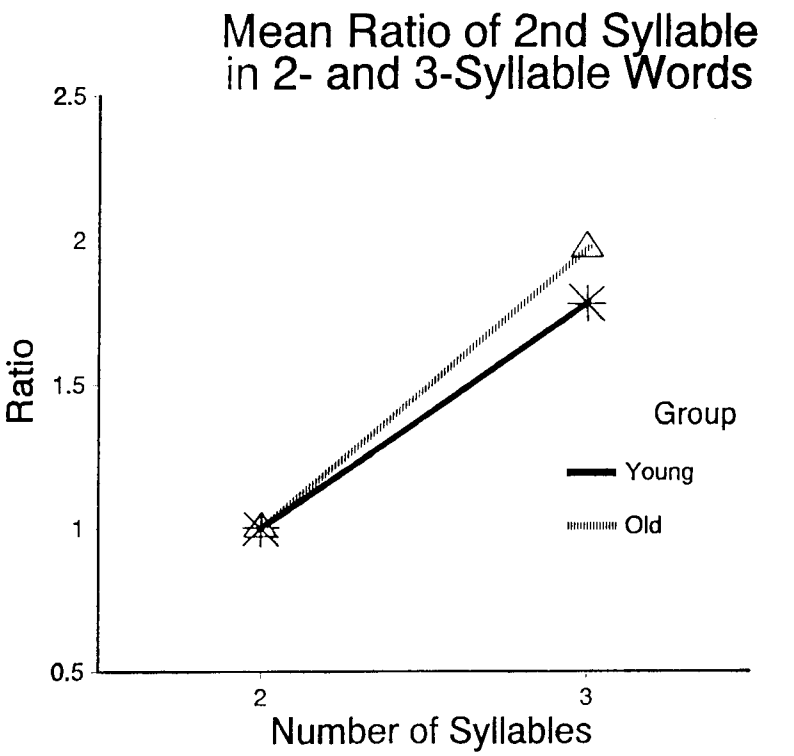

Figure 5. Mean ratio of 2nd syllable in bisyllabic words to 2nd syllable in trisyllabic words for each group. Ratio of bisyllabic duration to bisyllabic duration is equal to 1.0.

Discussion

The findings herein clearly demonstrate the effects of stress on syllable timing at the word level. Stressed syllables are longer than unstressed syllables. At a casual or moderately fast speaking rate (cf. "allegretto", Hiranburana, 1971), rhythmic patterns for the bisyllabic and trisyllabic compounds are iambic and anapaestic, respectively. These rhythmic patterns are present in the speech of Thai adults whether viewed by an absolute or relative standard of measurement. It, of course, is not impossible that changes in intensity, spectrum, and fundamental frequency also correlate with the distinction between stressed and unstressed syllables. At the word level, however, it would appear that duration is the dominant cue.

The findings in the present study are in agreement with earlier studies which have presented timing information for polysyllabic words, especially Luangthongkum (1977). Her data for a single speaker similarly point to iambic and anapaestic rhythmic patterns for bisyllabic and trisyllabic compounds, respectively, at a moderately fast or casual speaking rate. She concludes that syllable duration is the primary acoustic correlate of Thai stress. The findings not only provide empirical support for this notion, but also establish the magnitude of durational differences as a function of stress in polysyllabic words. In this study, speaking rate was held constant. Further large-scale studies on this topic are warranted to determine the relationship between the acoustic correlates of stress and speaking rate in Thai (cf. Hiranburana, 1971).

The fact that old speakers were more variable than young speakers on monosyllabic and trisyllabic words is probably indicative of a less efficient nervous system. Though as Smith et al. (1987) point out, "such differences are likely to be the complex result of various factors associated with sensory, motor, biomechanical, cognitive, linguistic, and other changes that occur in conjunction with normal aging" (p. 256). Bisyllabic words were similarly more variable for old speakers, and the group effect just narrowly failed to reach statistical significance at the .05 level. For young and old speakers alike, monosyllabic words were more variable in production than either bisyllabic or trisyllabic words. This suggests that variability is not merely a function of duration. Polysyllabic words may be less variable because of rhythmic constraints imposed by the stress patterns. There is less latitude in polysyllabic words if one is to produce appropriate timing patterns. In contrast, variability in production of voicing of Thai word-initial stops does not differ significantly as a function of age (Gandour et al., in press). Evidently, variability in speech timing increases more rapidly with normal aging at the syllable/word level than at the segmental level.

The average size of the difference in word duration between groups in this study (7%) is small by comparison to the study by Smith et al. (1987). They found that segment, syllable, and sentence durations of elderly adults were 20 to 25% longer than those of young adults. However, their aged group was considerably older than the one in this study. Their elderly adults ranged from 66 to 75 years of age. Older subjects in this study were between 53 and 60 years of age. Linguistic and methodological differences between the two studies notwithstanding, an increase in word duration is to be expected in a comparable sample of elderly Thai adults.

In conclusion, the present findings provide quantitative support for duration as a primary acoustic correlate of Thai stress. The rhythmic patterns that result from compound words reflect those that have been postulated in earlier phonological analyses. Such a database can serve as a springboard to future investigations of interword as well as intraword timing relations in Thai.

Notes

This material is based upon work supported in part by the National Institute on Deafness and other Communication Disorders under Grant No. DC00515-05 and the United States Information Agency in conjunction with the Thailand-United States Educational Foundation under a Fulbright Senior Researcher Award (1st author). The speech data were collected in Bangkok with the approval of the National Research Council of Thailand during the period from December 1988 to July 1989. In Bangkok, the first author was affiliated with the Division of Communication Disorders, Faculty of Medicine, Mahidol University, and Prasart Neurological Hospital. We are grateful to Andrea Billey and Siripong Potisuk for assistance with data processing.

References

Allen, G. (1973). Segmental timing control in speech production. *Journal of Phonetics*, 1, 207-225.

Bee, P. (1975). Restricted phonology in certain Thai linker syllables. In J. Harris & J. Chamberlain (Eds.), *Studies in Tai linguistics in honor of William J. Gedney*. Bangkok : Central Institute of English Language.

Collins, M., Rosenbek, J., & Wertz, R. (1983). Spectrographic analysis of vowel and word duration in apraxia of speech. *Journal of Speech and Hearing Research*, 26, 217-224.

Gandour, J., Holasuit Petty, S., & Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language, 35*, 201-240.

Gandour, J., Ponglorpisit, S., & Dechongkit, S. (in press). Age-related effects on the production of voice onset time in Thai word-initial stops. *Pasaa*, 20(2), 33-39.

Gandour, J., Ponglorpisit, S., Khunadorn, F., Dechongkit, S., Boongird, P., & Boonklam, R. (1992). Timing characteristics of speech after brain damage: Vowel length in Thai. *Brain and Language, 42*, 337-345.

Hiranburana, S. (1971). *The role of accent in Thai grammar*. Unpublished doctoral dissertation, University of London, School of Oriental and African Studies.

Kent, R., & Forner, L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics, 8*, 157-168.

Klatt, D. (1976). Linguistic uses of segmental duration in English : Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*, 1208-1221.

Luangthongkum, T. (1977). *Rhythm in standard Thai*. Unpublished doctoral dissertation, University of Edinburgh.

Luksaneeyanawin, S. (1983). *Intonation in Thai*. Unpublished doctoral dissertation, University of Edinburgh.

Noss, R. (1972). Rhythm in Thai. In J. Harris & R. Noss (Eds.), *Tai phonetics and phonology* (pp. 33-42). Bangkok: Central Institute of English Language.

Peterson, G., & Lehiste, I. (1960). Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America, 32*, 693-703.

Rudaravanija, P. (1965). *An analysis of the elements in Thai that correspond to the basic intonation patterns of English*. Unpublished doctoral dissertation, Columbia University.

Smith, B., Sugarman, M., & Long, S. (1983). Experimental manipulation of speaking rate for studying temporal variability in Children's speech. *Journal of the Acoustical Society of America, 74*, 744-749.

Smith, B., Wasowicz, J., & Preston, J. (1987). Temporal characteristics of the speech of normal elderly adults. *Journal of Speech and Hearing Research, 30*, 522-529.

Thawisomboon, S. (1955). *Syllable junctures within stress groups in spoken Thai*. Unpublished Master's thesis, University of London, School of Oriental and African Studies.

Weismer, G., & Fennell, A. (1985). Constancy of (acoustic) relative timing measures in phrase-level utterances. *Journal of the Acoustical Society of America, 78*, 49-57.

Winer, B., Brown, D., & Michels, K. (1991). *Statistical principles in experimental design* (3rd ed.). New York : McGraw-Hill.

Appendix

Duration Measures on an Absolute Scale

Young Speakers (n=10)

Words[a]	S1[b]	S2	S3	Total
l	448 (69)	-	-	448 (69)
w	437 (81)	-	-	437 (81)
s	518 (88)	-	-	518 (88)
d	445 (101)	-	-	445 (101)
n	494 (89)	-	-	494 (89)
lk	255 (39)	473 (93)	-	729 (117)
wk	291 (42)	454 (90)	-	745 (122)
sd	273 (68)	511 (92)	-	783 (142)
ds	216 (47)	504 (69)	-	720 (103)
nt	317 (63)	443 (85)	-	760 (139)
lkb	216 (41)	219 (43)	455 (83)	882 (141)
wkp	232 (39)	270 (39)	305 (45)	808 (105)
sdk	235 (62)	317 (46)	382 (73)	935 (158)
dsp	190 (53)	306 (43)	435 (83)	932 (153)
ntt	254 (60)	246 (41)	454 (79)	955 (159)

Old Speakers (n = 10)

Words	S1	S2	S3	Total
l	448 (63)	-	-	448 (63)
w	463 (85)	-	-	463 (85)
s	540 (76)	-	-	540 (76)
d	469 (75)	-	-	469 (75)
n	526 (74)	-	-	526 (74)
lk	267 (51)	530 (64)	-	796 (100)
wk	306 (54)	518 (84)	-	824 (111)
sd	262 (57)	551 (79)	-	814 (111)
ds	248 (47)	532 (92)	-	785 (99)
nt	347 (60)	496 (92)	-	844 (131)
lkb	237 (52)	217 (43)	511 (67)	965 (126)
wkp	242 (45)	277 (44)	312 (40)	833 (108)
sdk	252 (75)	325 (51)	413 (70)	991 (146)
dsp	227 (53)	303 (63)	496 (95)	1025 (156)
ntt	286 (49)	257 (50)	492 (70)	1034 (136)

[a]Abbreviated labels for each of the 15 words. Letters represent the syllable-initial consonants. See also Table 1.

[b]S = syllable.

Note. Values are means (standard deviations) expressed in ms.

Duration Measures on an Relative Scale

Young Speakers (n = 10)

Words[a]	S1[b]	S2	S3
lk	35.3 (4.1)	64.6 (4.1)	-
wk	39.4 (3.4)	60.6 (3.4)	-
sd	34.6 (4.9)	65.4 (4.9)	-
ds	29.9 (3.7)	70.3 (3.9)	-
nt	41.6 (3.5)	58.3 (3.5)	-
lkb	24.7 (3.8)	24.8 (2.4)	50.3 (3.6)
wkp	28.7 (3.1)	33.5 (2.6)	37.8 (2.4)
sdk	25.0 (4.0)	34.2 (3.1)	40.8 (3.2)
dsp	20.2 (3.1)	33.1 (3.4)	46.7 (4.3)
ntt	26.4 (2.8)	25.9 (2.9)	47.7 (3.8)

Old Speakers (n = 10)

lk	33.3 (3.7)	66.7 (3.7)	-
wk	37.2 (4,6)	62.8 (4.5)	-
sd	32.1 (4,7)	67.8 (4.8)	-
ds	31.6 (4.7)	67.6 (7.2)	-
nt	41.2 (4.2)	58.7 (4.4)	-
lkb	24.4 (3.1)	22.5 (3.0)	53.1 (4.6)
wkp	29.0 (2.7)	33.2 (2.7)	37.5 (3.5)
sdk	25.1 (4.6)	32.9 (3.3)	41.9 (5.5)
dsp	22.3 (4.8)	29.5 (3.6)	48.3 (4.8)
ntt	27.6 (3.1)	24.8 (2.8)	47.6 (4.0)

[a]Abbreviated labels for each of the 15 words. Letters represent the syllable-initial consonants. See also Table 1.

[b]S = syllable.

Note. Values are means (standard deviations) expressed in percent.