LINGUISTICS OF THE TIBETO-BURMAN AREA
VOLUME 17.2 — FALL 1994

# $F_0$ correlates of stress in Thai*

**Siripong Potisuk**
School of Electrical Engineering, Purdue University
**Jackson T. Gandour**
Department of Audiology and Speech Sciences, Purdue University
**Mary P. Harper**
School of Electrical Engineering, Purdue University

*Abstract*

An experiment was conducted to investigate changes in the fundamental frequency ($F_0$) contours of Thai tones in connected speech as a function of stress. Thai has five tones: mid, low, falling, high, and rising. Stimuli consisted of 25 pairs of ambiguous target sentences with disambiguating context, produced at a conversational speaking rate. One member of each pair contained a 2-syllable noun-verb sequence exhibiting a - - stress pattern, the other member a 2-syllable noun compound exhibiting a ˘ - stress pattern. Acoustic analysis revealed that $F_0$ contours of stressed syllables more closely approximate $F_0$ contours in citation forms than those of unstressed syllables. The degree of approximation is primarily determined by syllable structure and the interaction between adjacent tones. In contrast, $F_0$ contours of unstressed syllables undergo a more complex process. The average height of all five tones can be classified into three tonal registers: low, mid, and high. The low register comprises the low and the rising tones, the mid register the mid tone, and the high register the falling and the high tones. Based on shape, the falling and high tones are distinguished within the high register, the low and rising tones within the low register. Therefore, a five-way contrast among all five tones appears to be maintained in both stressed and unstressed syllables. In addition, two statistical parameters, average $F_0$ and coefficient of variation, are proposed for a machine model to automatically detect stressed and unstressed syllables.

* Acknowledgments: This material is based upon work supported in part by the National Institute on Deafness and other Communication Disorders under Grant No. DC00515-07, and the United States Information Agency in conjunction with the Thailand-United States Educational Foundation under a Fulbright Senior Researcher Award (2nd author). In Thailand, the first author, who is currently on a leave of absence from the Academic Division of Chulachomklao Royal Military Academy, extends his gratitude to the Royal Thai Army. We also wish to thank Tom Quesnell for his assistance with $F_0$ 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.

## Introduction

The impetus for this research arose during an investigation on automatic detection of stress in Thai by computer—part of an ongoing research project on automatic speech understanding of Thai at Purdue University. It is of primary interest to be able to formulate an algorithm that will extract and utilize acoustic correlates of stress such as fundamental frequency ($F_0$), duration, and intensity of the speech waveform to identify the location of stressed and unstressed syllables and, if possible, to assign them to their appropriate tonal categories. Thus, it is necessary to develop quantitative measures that will enable computers to accurately perform the task.

Stress in Thai has been investigated in the past by many linguists (Henderson, 1949; Noss, 1964; Warotamasikkhadit, 1967; Hiranburana, 1971; Surintramont, 1973; Gandour,1975; Luangthongkam, 1977; Sutadarat, 1978; Luksaneeyanawin, 1983), and many stress placement rules have been postulated. In terms of pitch, those researchers generally agreed on the phonetic realization of stress in "linker syllables" (Bee, 1975). However, the effects of stress on $F_0$ contours of unstressed non-linker syllables remain a subject of much controversy up to the present. The disagreement revolves around the issue of whether or not lexical tones of unstressed syllables undergo tone neutralization, i.e., whether or not $F_0$ contours of all or some of the five tones lose their identities in both height and shape. Of those earlier studies, only a few have presented acoustic-phonetic information on the realization of stress in terms of $F_0$.

Hiranburana (1971, 1972) presented instrumental findings on stress at the word level. Changes in $F_0$ contours were shown to vary depending on degree of stress. Her results were based on observations of the pattern of changes in $F_0$ contours of unstressed syllables obtained from non-final syllables of polysyllabic words, monosyllabic grammatical words, the first sylllable of institutionalized compounds, and the reduplicator of the completely reduplicative forms. She concluded that $F_0$ contours of the five lexical tones are neutralized to three level tones: high, mid, and modified low.

Gandour (1975) argued against tone neutralization in fast casual speech by presenting acoustical measurements of $F_0$ contours of the initial syllable in pairs of bisyllabic noun compounds distinguished minimally or near-minimally by the lexical tone of the initial syllable. His findings indicated no changes in contour tones of unstressed syllables. He concluded that the five lexical tones of unstressed syllables maintain their basic canonical shapes as in citation forms, despite being shorter in duration, so that the five-way contrast is intact.

Luksaneeyanawin (1983) extended the phonetic analysis of stress beyond the word level. Based on acoustic and auditory analyses of passages read by two speakers, her descriptions of $F_0$ contours of unstressed non-linker syllables were generally in agreement with Hiranburana's, except those of the

rising tone. It appears that she also favors the existence of tone neutralization in unstressed non-linker syllables.

The findings in all three aforementioned studies are unfortunately very difficult to interpret because none of them isolated stress from other confounding factors affecting the realization of $F_0$ contours, such as tonal coarticulation, declination and intonation. It is well known that tonal coarticulation or the interaction between adjacent tones plays a major role in determining the height and shape of $F_0$ contours of tones (Shen 1990). Therefore, their results might not be due to the manifestation of stress alone, but to other factors as well.

The aim of this study is to investigate changes in $F_0$ contours of the five Thai tones in connected speech as a function of stress by using a more systematically controlled experiment so that the effects of stress may be accurately assessed and quantified. The focus will primarily be on the effects of stress occurring in bisyllabic noun compounds. Compounds in Thai are very important not only because of their high frequency of occurrence, but because they provide us with a window to see how prosody may potentially be used by listeners to resolve ambiguities in Thai. The study will attempt to answer questions concerning the effects of stress on individual tones and the constrastive relationship of lexical tones in both stressed and unstressed syllables. Findings will be interpreted in terms of their relevance to the description of sentence prosody in Thai. Implementation issues regarding automatic stress detection will also be addressed.

## Method

*Subjects*

Three native speakers of Thai at Purdue University participated in this preliminary study. CT, male, 22 years old, was an undergraduate student; KW and SU, female, 25 and 32 years old, respectively, were graduate students. All three subjects were native speakers of Bangkok Thai dialect and were naive with respect to the purpose of the experiment.

*Materials*

Stimuli consisted of 25 pairs of ambiguous target sentences. The two members of each pair contained six segmentally identical syllables including two target syllables. The first member (a) contained a 2-syllable noun-verb sequence exhibiting a - - stress pattern, the second member (b) a 2-syllable noun compound exhibiting a ˘ - stress pattern. The diacritic ˘ represents an unstressed syllable, - a stressed syllable. To minimize tonal coarticulation

effects, the two target syllables were embedded at the beginning of the sentence, hence only anticipatory coarticulation on the first syllable is present, while carryover coarticulation is eliminated. Thai tones are more greatly influenced by carryover than anticipatory coarticulations (Gandour et al., in press). The tones of the two target syllables were also varied to represent all possible two-tone combinations of five Thai tones so that anticipatory coarticulations in all contexts are considered. Of 25 two-tone combinations, only four were fully voiced throughout (MH, MR, LF, and FH); the other 21 two-tone combinations had intervening voiceless obstruents. To maximize the speaker's likelihood of being able to naturally produce the utterence according to its intended meaning, each utterance was preceded by a few sentences of disambiguating context. A list of the target sentences with their disambiguating contexts is included in the Appendix.

*Recording Procedure*

Speakers were asked to read a target sentence along with a few sentences of disambiguating context typed in Thai script on a 5 x 8 in. card. Cards were presented in random order and speakers were not told which of the sentences in the paragraph was the target sentence. They were also instructed to produce the sentences at a conversational speaking rate, i.e., at a rate they considered representative of their conversational speech. A random order of presentation and a sufficient pause provided between items were intended to minimize changes in speaking rate and learning or list-reading effects, thus maximizing the likelihood of speakers being able to produce natural sounding utterances. To avoid start and end effects, extra cards were placed at the top and bottom of the deck.

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. For each speaker, the total corpus contained 250 utterances (2 members x 25 tonal combinations x 5 repetitions). There were two recording sessions separated by one week to minimize the possibility of speakers' exaggerating the contrast between the two members (Price et al.,1990). The (a) members of all pairs were assigned to the first recording session, the (b) members to the second session. Before the recording session began, the speakers were allowed to familiarize themselves with the target sentences. During the session, speakers were asked to reread any sentences that the investigators deemed "off-target" until an acceptable version was produced. Each session lasted about 45 minutes.