

In Tai Linguistics in Honor of
Fang-kuei Li, ed. by Thomas W.
 Getthing, Jimmy G. Harris, and

THAI TONES AS A REFERENCE SYSTEM

Pranee Kullavanijaya.

Arthur S. Abramson

Bangkok: Chulalongkorn

The University of Connecticut and Haskins Laboratories

University Press

1976.

In a tone language the question arises as to how much information is needed by the listener to identify words minimally differentiated by phonologically relevant tones. Relative, rather than absolute, values of the fundamental frequency (f_0) of the voice are normally found to provide the major acoustic cues for the identification of phonemic tones even in the presence of other phonetic features.¹ Certain tones of a given system may be quite identifiable in isolated citation forms of syllables, while others may have to be embedded in a linguistic context for easy recognition. Even the latter tones, however, may enjoy high identifiability in isolation if the listener is free to adapt his perception to the tone system of a given speaker. The perceptual assessment of the tones of a language may well be analogous to that of a vowel space. In the present study² experiments were designed to test the hypothesis that in one tone language, Central Thai (Siamese), listeners would do less well in identifying the tones when deprived of an opportunity to become oriented to the tonal space of the speaker.

Thai is traditionally viewed as having five phonemic tones.³ The maximum five-way differentiation can appear only on syllables ending in diphthongs, long vowels, or nasal consonants.⁴ In principle then, such a syllable can manifest itself as five different monosyllabic words or morphemes, each with its own tone. Every item in the Thai lexicon is characterized by phonologically relevant tone as well as by consonantal and vocalic features. The five tones can be divided into two groups, the dynamic tones and the static tones (Abramson 1962: 9). In this scheme, the rather sharp downward f_0 movement of the falling tone and upward movement of the rising tone place them in the dynamic category. It is their movements rather than their end points that seem to characterize them. Since the high, mid, and low tones have sounded to many observers as if they simply occupy three pitch levels, they are classified as static; nevertheless, these tones too, especially the high and low, also show some movement although it tends to be fairly slow (Abramson 1962:120-7; Erickson 1974).⁵ This will be relevant to the discussion of the results of the experiments to be presented here. If the static/dynamic dichotomy is perceptually valid, we might expect the static tones to be more readily confused under difficult listening conditions than the dynamic tones.

Procedure

The following Thai monosyllabic words minimally differentiated by tone were recorded by ten native speakers of Central Thai, five men and five women:

Tone	Word	Gloss
Mid	/khaa/	'a grass (<i>Imperata cylindrica</i>)'
Low	/khàa/	'galangal, a rhizome'
Falling	/khâa/	'slave, servant'
High	/kháa/	'to engage in trade'
Rising	/khǎa/	'leg'

These tape recordings were randomized into composite tests and individual tests. In the composite tests the ten speakers and all their productions were randomized to make it impossible for the listener to predict at each moment not only what word was going to be said but also which of the ten voices was going to say it. In each of the individual tests the speaker remained the same throughout; the only uncertainty for the listener was the identity of each word as it was played into his headphones. Responses written in Thai script were provided by thirty four native speakers over a period of approximately one month. The composite test, in two randomizations, was administered first. In each of the test orders there were three tokens of each word for each of the ten speakers. Each of the ten individual tests was prepared in four randomizations with five tokens of each word in each test order. Since there was not enough time to administer more than one test order for each individual test, care was taken to ensure that no single randomization was presented to the subjects more often than once a week.⁶ All the recordings were checked for acceptability by the speakers themselves and the experimenter.

Results

The responses of all 34 subjects to the composite tests are displayed in percentages in the form of a confusion matrix in Table I.

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	82.4	17.0	0.3	0.3	
Low	6.9	92.9	0.1	0.1	0.1
Falling	0.2	0.7	98.9		0.2
High	0.3	0.1	0.3	98.7	0.6
Rising	0.1		0.3	0.1	99.6

Responses = 8847

Subjects = 34

% Correct = 94.4

Table I

Composite tests. Pooled data for 10 speakers

The stimuli are listed in the first column of the matrix and the response labels are arranged across the top. Correct responses to the stimuli as intended are entered in the cells along the diagonal from the upper left to the lower right⁷. An overall percent correct of 94.4 seems rather high until one observes in the confusion matrix that nearly all the errors are caused by confusion between only two of the tones, the mid and low. The mid tone is heard as low 17% of the time, and the low tone is heard as mid 6.9% of the time. These confusions are largely caused by four of the speakers whose data will be displayed separately. In addition, about half of the test subjects are responsible for most of the confusions. The responses to the ten separate individual tests are pooled in Table II.

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	99.2	0.7		0.1	
Low	0.9	99.0		0.1	
Falling		0.1	99.5	0.2	0.1
High	0.1		0.2	99.5	0.1
Rising	0.1		0.2		99.7

Responses = 7450

Subjects = 34

% Correct = 99.4

Table II

Individual tests. Pooled data for 10 speakers

We see an overall improvement of 5%, but what is more important is the virtual elimination of the confusions between the mid and low tones. There is also a smaller scattering of errors over the rest of the matrix.

The major results of the experiment are seen more clearly if we examine the responses to the productions of the four people responsible for most of the errors. The selection criterion was a score of 75% or less in any cell of a speaker's confusion matrix for the composite test. This threshold was based on the following reasoning. The chance level for each stimulus is nominally 20%, since there was a five-way choice; however, the single serious confusion, that between the low and mid tones, yields in fact a chance level of 50%. The threshold chosen was halfway between 50% and 100%; thus three speakers had a score of 75% or less for the mid tone and one speaker for the low tone in the composite test.

In Tables III and IV we see the composite and individual data respectively for Speaker A, a women.

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	66.1	33.3		0.6	
Low	0.6	99.4			
Falling		0.6	99.4		
High			0.6	96.0	3.4
Rising				1.1	98.9

Stimuli

Responses = 885 Subjects = 34 % Correct = 92.1

Table III
Speaker A's composite data

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	98.0	2.0			
Low	0.7	99.3			
Falling			100		
High				100	
Rising					100

Responses = 750

Subjects = 30

% Correct = 99.5

Table IV

Speaker A's individual data

Table IV shows an overall improvement of 7.4%, mostly to be attributed to the virtual elimination of a third of the responses to the mid tone as low in the composite test. In addition, the 3.4% identification of high as rising also disappears.

The data for Speaker C, a man, appear in Tables V and VI.

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	39.0	58.7	2.3		
Low	3.4	96.6			
Falling		0.6	99.4		
High	0.6			99.4	
Rising					100

Responses = 885

Subjects = 34

% Correct = 90.9

Table V

Speaker C's composite data

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	99.2			0.8	
Low	5.4	93.9		0.8	
Falling		1.5	97.7	0.8	
High	1.5			97.7	0.8
Rising		0.8	0.8		97.7

Responses = 650

Subjects = 26

% Correct = 97.4

Table VI
Speaker C's individual data

The effect is even more striking here in that more than half of the responses to the mid tone in the composite test are called low, while there are no responses to it as low in the individual test. It is true, however, that the low tone itself drops a few percentage points in the individual test with the errors assigned to the mid tone; that is, this small confusion in the composite test is not removed in the individual test.

The data for Speaker F, a man, are shown in Tables VII and VIII.

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	75.1	24.9			
Low	2.3	97.7			
Falling	0.6	2.3	96.6		0.6
High			2.3	97.7	
Rising					100

Responses = 885

Subjects = 34

% Correct = 93.5

Table VII
Speaker F's composite data

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	100				
Low		100			
Falling			100		
High				100	
Rising					100

Responses = 750

Subjects = 30

% Correct = 100

Table VIII

Speaker F's individual data

They conform to the general trend but in a more satisfyingly clear-cut manner since all scores become 100% along the diagonal in Table VIII.

The data for Speaker J, a man, shown in Tables IX and X,

% Responses

Labels:	Mid	Low	Falling	High	Rising
Mid	100				
Low	35.4	64.6			
Falling	0.6		98.9		0.6
High				100	
Rising	0.6				99.4

Responses = 885

Subjects = 34

% Correct = 92.7

Table IX

Speaker J's composite data

Labels:	Mid	Low	Falling	High	Rising
Mid	100				
Low	0.7	99.4			
Falling			100		
High				100	
Rising					100

Responses = 775

Subjects = 31

% Correct = 99.9

Table X
Speaker J's individual data

support the hypothesis stated at the outset but do so in a different way. It is not the mid tone which tends to be misheard in the composite test but rather the low tone, which is heard as mid 35.4% of the time.

For the other six speakers the general pattern is the same, although the effects are somewhat smaller. The mid tones of five of them are sometimes heard as low in the composite test while the low tones of the sixth person are sometimes heard as mid. These effects disappear in the individual tests.

The thirty-four listeners were not equally confused by the utterances of the mid and low tones produced by the four speakers of Tables III - X. Every subject made at least some errors, but some listeners were much more affected than others. In addition, the subjects varied somewhat as to which speaker confused them the most. Every one of the four speakers had 100% correct responses to the two tones in question for at least a few listeners. For example, Speaker A caused no errors in the responses of eight subjects to the composite test. Even Speaker C, whose mid tone was heard as low 58.7% of the time, had three listeners who made no errors in labeling his productions of the two tones.

Discussion

We may conclude that in Thai, and probably in other tone languages, phonemic tones free of a linguistic context are better iden-

tified when the listener has access to the speaker's tone space. The two Thai tones that are most vulnerable to confusion are the mid and low tones, which are typically characterized by very little movement of the fundamental frequency of the voice over time. This can be seen in Figure 1 for one of the speakers least productive of errors in this study, i.e. not one of the four singled out for special attention. The figure shows Speaker P.C.'s f_0 curves normalized for time and stated as percentages of the total voice range used by her for the test items. The falling, rising, and high tones--thus two dynamic tones and one static tone--appear to have such distinctive contours that their great invulnerability to confusion in the composite condition is not at all surprising. Indeed, the high tone of citation forms typically includes quite audible glottal constriction or creak.

The question still arises as to why it is that the mid tone is normally confused with the low tone rather than the opposite in the composite condition.⁸ Figure 1 shows that the low tone, as compared with the mid tone, drops fairly abruptly and then descends slowly toward the bottom of the voice range. When this characteristic movement is very evident to the listener, it may render the low tone robust. Such a contour would seem to be a sufficient cue for the low tone but not a necessary one, because the down drift of the mid tone in the confusion of the composite condition may be enough to make some listeners uncertain and cause them to assign it to the only possible other choice, namely the low tone.

The opposite effect for Speaker J (Table IX), labeling the intended low tone as mid, is more difficult to explain. It should be noted that when a speaker's low tone reaches the bottom of his voice range quite early and stays there until the end, vocal fry can often be heard in citation forms. The low end of Speaker J's voice range is defined by the bottom point of his rising tone and not by his low tone. (For Speaker P.C. in Figure 1 the bottom of her range is reached by both these tones as well as the falling tone.) We may speculate then that Speaker J's mid-tone contour is robust in that its shape cannot be mistaken for anything else even in the composite condition, while his low tone simply does not possess the voice quality often associated with that tone when it clings for a good part of its duration to the bottom of the voice range.

Given the disagreements in the literature over the stability of this tonal opposition (Noss 1954, 1964; Abramson 1962, 1972, 1975), it is interesting to find that this contrast is so vulnerable to confusion in the composite condition. Richard B. Noss (1975:279) comments, "It would appear that, while individual speakers keep the mid and low tone contours distinct, there may be little or no differentiation of these contours between speakers." His view, although somewhat overstated, receives support from a comparison of the composite and individual conditions of the present study.

CITATION FORMS

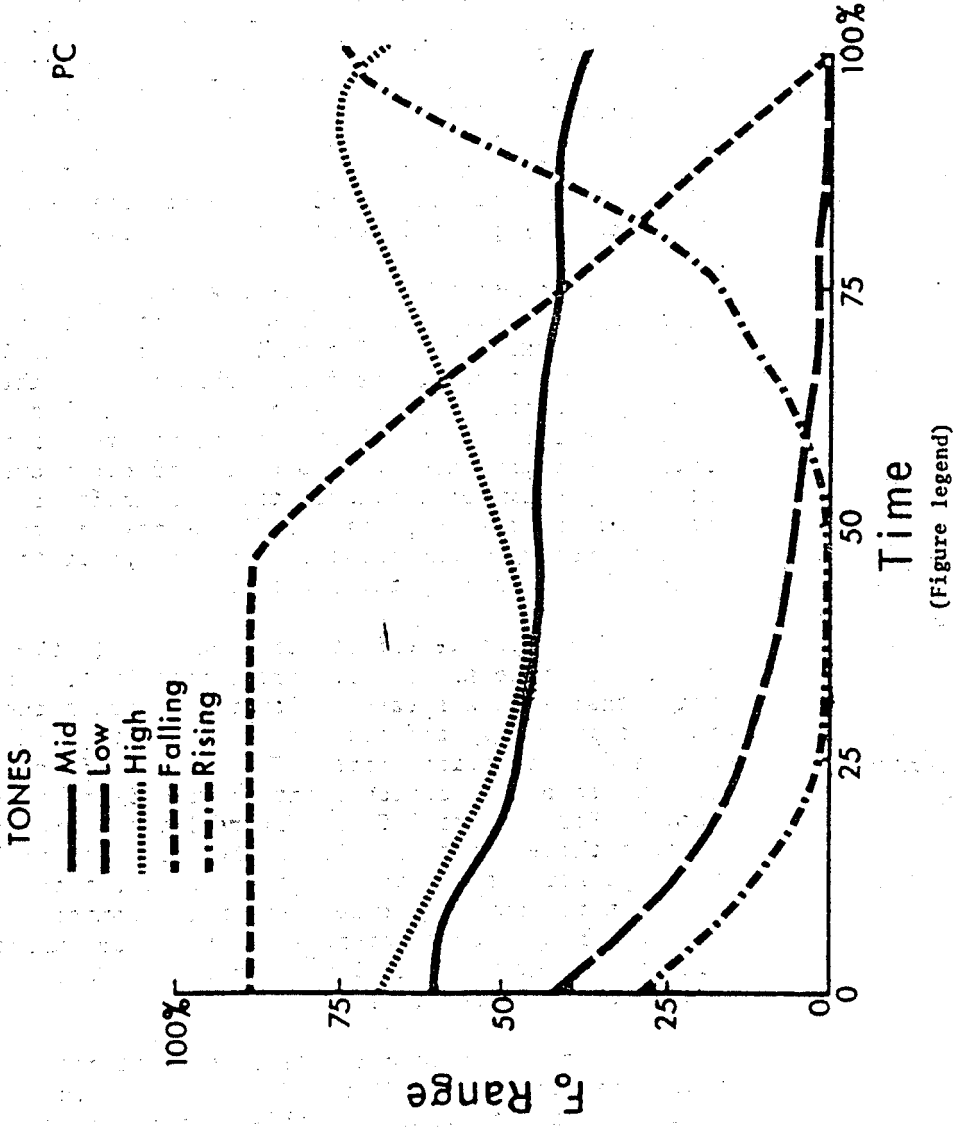


Figure 1. Average f_0 contours of speaker P.C. produced for the tests.

Finally, the individual differences observed in production and perception lead to one other generalization. Some Thai speakers seem to provide minimal cues for the distinction between the mid and low tones. Many listeners require more than these minimal cues if they are to identify the tones of those speakers when no opportunity is given to become accommodated to the tonal space of each such speaker.

NOTES

- 1 E.g., creaky voice, voice breaking, and amplitude variations.
- 2 While most of the analysis of data was performed at Haskins Laboratories, the data themselves were collected while the author was on sabbatical leave in Thailand on research fellowships from the American Council of Learned Societies and the Ford Foundation Southeast Asia Fellowship Program. I gratefully acknowledge the hospitality of Dr. Udom Warotamasikkhadit, Dean, the Faculty of Humanities, Ramkhamhaeng University and Mrs. Mayuri Sukwiwat, Director, the Central Institute of English Language, both in Bangkok. An oral version of this paper was presented at the Fifth Essex Symposium on Phonetics, August 25-27, 1975, University of Essex, Colchester, England.
- 3 Here I shall not enter into the question of whether the five tones can be profitably analyzed into some smaller number of underlying distinctive features. This question has recently been discussed in Gandour 1975.
- 4 Other syllable types support fewer than five tones.
- 5 Without undue strain one can hear pitch glides in the high and low tones. Indeed, f_0 contours of the five tones in running speech (Abramson Ms) render the dynamic/static dichotomy unclear; however, it is mentioned here because of its apparent auditory usefulness.
- 6 The tests were conducted in the language laboratory of Ramkhamhaeng University, Bangkok, Thailand. The subjects, all undergraduate students, were chosen and instructed by Miss Phaanit Chotibut of the Department of English; her enthusiastic help is much appreciated.
- 7 Variations in the number of responses and the number of subjects in the tables to follow indicate fluctuations in attendance at the test sessions and occasional omissions on the answer sheets.
- 8 The data do not support any influence of sex.

REFERENCES

- Abramson, Arthur S. 1962. "The vowels and tones of standard Thai: Acoustical measurements and experiments." *Indiana University Research Center in Anthropology, Folklore, and Linguistics* 20.
- _____. 1972. "Tonal experiments with whispered Thai." *Papers in linguistics and phonetics to the memory of Pierre Delattre*, ed. by A. Valdman. The Hague: Mouton.
- _____. 1975. "The tones of Central Thai: some perceptual experiments." *Studies in Tai linguistics in honor of William J. Gedney*, ed. by Jimmy G. Harris and James R. Chamberlain. Bangkok: Central Institute of English Language.
- _____. Ms. 1975. "The coarticulation of tones: an acoustic study of Thai." *Proceedings of the Eighth International Congress of Phonetic Sciences*.
- Erickson, Donna. 1974. "Fundamental frequency contours of the tones of standard Thai." *Pasaa: Notes and News about Language Teaching and Linguistics in Thailand* 4: 1-25.
- Gandour, Jack. 1975. "On the representation of tone in Siamese." *Studies in Tai linguistics in honor of William J. Gedney*, ed. by Jimmy G. Harris and James R. Chamberlain. Bangkok: Central Institute of English Language.
- Noss, Richard B. 1954. *An outline of Siamese grammar*. Yale University dissertation.
- _____. 1964. *Thai reference grammar*. Washington, D.C.: Foreign Service Institute.
- _____. 1975. "How useful are citation forms in synchronic Thai phonology?" *Studies in Tai linguistics in honor of William J. Gedney*, ed. by Jimmy G. Harris and James R. Chamberlain. Bangkok: Central Institute of English Language.