PART III

International Journal of American Linguistics

Volume 28 Number 2
APRIL 1962

Publication Twenty
of the
INDIANA UNIVERSITY RESEARCH CENTER
in
ANTHROPOLOGY, FOLKLORE, AND LINGUISTICS

THE VOWELS AND TONES OF STANDARD THAI: ACOUSTICAL MEASUREMENTS AND EXPERIMENTS

by

Arthur S. Abramson

Second-class postage paid at Baltimore 2, Md.

THE VOWELS AND TONES OF STANDARD THAI: ACOUSTICAL MEASUREMENTS AND EXPERIMENTS

THE VOWELS AND TONES OF STANDARD THAI:

ACOUSTICAL MEASUREMENTS AND EXPERIMENTS

Arthur S. Abramson

Bloomington, 1962

INDIANA UNIVERSITY
RESEARCH CENTER IN
ANTHROPOLOGY, FOLKLORE, AND LINGUISTICS

Editor: C. F. Voegelin

Director of Publications: Thomas A. Sebeok

Publication Twenty, pp. x + 146, Issued April 1962
Also
Part II of the
International Journal of American Linguistics
Vol. 28, No. 2 (1962)

Library of Congress Catalog Card Number: 62-62698

Priced separately, $3.00. Orders will be received by the
Director of Publications of the Research Center, at
Indiana University, Bloomington, Indiana

PREFACE

This study is a slightly revised version of a Columbia University Ph. D. dissertation that appeared under the same title in June 1960 (University Microfilms, Ann Arbor, Michigan). With the help of the Haskins Laboratories I was able to distribute copies of my dissertation to a number of people in the field. Many of them were kind enough to comment on one aspect or another of the study. I am especially grateful to Professor George L. Trager for going to the trouble of writing a lengthy criticism of the section on the Phonological Structure.

For the most part, these critical comments are tangential to the main body of the work, but in preparing the present edition, wherever I have seen my way clear to taking them into account, I have tried to do so in footnotes. In addition, I have made such changes in the text as seemed likely to contribute to greater cohesiveness and clarity. In the Summary and Outlook I have added a few passages on new work, mostly still in progress, that is a continuation of the experiments reported here.

The research for this study was done during the period of 1957-1960 in the Department of Linguistics of Columbia University and the Haskins Laboratories in New York City. I benefited considerably in my work from my extensive contact with the Thai language in Thailand during my tenure as a Fulbright teacher of English from April 1953 to July 1955.

I wish to thank my advisor, Professor John Lotz of Columbia University, not only for the important part he played in my linguistic training but also for the meticulous help and advice that he has always willingly extended to me in the conduct of my research. It was indeed he who urged me, early in my graduate studies, to study acoustic phonetics and perhaps combine an interest in it with other linguistic interests.

As my teacher and advisor, Professor Franklin S. Cooper of Columbia University and the Haskins Laboratories has been a constant source of both personal and scientific inspiration. His gentle but carefully thought out guidance has been indispensable to me in my research and experimentation.

The unique facilities of the Haskins Laboratories in New York City made the particular direction of my research possible. Both before and during my tenure on the research staff of the Laboratories, I have benefited in my work on Thai from my association with the happy blend of experimental psychologists, linguists, and communications engineers working together there. I further wish to acknowledge the contribution made by the Haskins Laboratories toward the costs of printing this study.

I am grateful to another of my teachers, Professor Joseph H. Greenberg of Columbia University, for making extensive access to informants and test subjects feasible.

The informants on whom I leaned most heavily in this country were Mr. Ekawit Nathalang and his brother Mr. Wathanyuu Nathalang, Miss Khomkhaay Chongcharoensuk, and Miss Kandaa Sindhvananda, all of the Ministry of Education, Bangkok, Thailand. I owe special thanks to Mrs. Patra Aphichaisiri, Principal of the Benchamaraajaalai School in Bangkok, who enthusiastically taught me the language during my stay in Thailand.

Final copies of all diagrams and other illustrations were executed for reproduction by Miss Agnes McKeon of the Haskins Laboratories. Her precise and attractive work has certainly contributed to whatever clarity the text may have.

CONTENTS

INTRODUCTION

A. Scope of the Study

The language under consideration is Standard Thai (Siamese). The name Siamese is better known in the West, although the name Thai has come into greater international use since the Second World War. In any event, Thai (/tʻaj/) is the name used in daily Thai speech. Standard Thai is the national language of Thailand and the dialect of the Central Region including the capital Bangkok. Except for its use in educational, official and commercial activities, other major Thai dialects prevail in the rest of the country.

The study begins with an outline of the phonology of Standard Thai. Although this analysis is largely based on field work done with informants[1] both in Thailand and the United States, it has obviously been influenced and guided by the work of others. Such previous work is acknowledged where relevant, but it is the work of Mary Haas (see Bibl.) that has been most influential. This brief statement of the phonology is presented in the Introduction not merely for its inherent interest but mainly as a linguistic framework for the acoustical investigation.

This research was designed to make a contribution to Thai linguistics through an acoustical and perceptual study of Thai vowels and tones. Physical data on these phonemes are made available by the techniques of acoustical analysis. The analysis is restricted for the most part to short utterances, although running speech is also considered to some extent. It is believed that the present study can serve as a foundation for work on consonantal features and on connected discourse. Complementary to the analytic aspects of the study are the experiments on the perception of the vocalic and tonal features that emerge from the acoustical measurements. These experiments were performed by means of speech synthesis and the manipulation of real speech. Suggestions for further work are made at various points.

It is hoped that, aside from such specialized interest as it may have, this study will also represent a modest contribution to the fast growing field of acoustic phonetics. It is at least an indication of the kind of research that can be done with a structural linguistic approach by phoneticians who find it desirable and convenient to avail themselves of existing devices and techniques and who may wish to help in the development of new technical approaches.

B. Phonological Structure

This treatment is not intended as an exhaustive description of the phonology of Standard Thai. It is, rather, an outline that is sufficiently detailed to serve as a framework within which to regard the acoustical measurements and experiments. Matters of juncture, stress, and intonation, to be discussed under Syntagmatic Phonology and Emphatic Features, have been analyzed in greater detail by others[2] with varying conclusions. Final solutions to these difficult problems are not crucial to the present work.

1. The Syllable

The distinctive phonemes are established within the syllabic frame. The phonemic syllable in Thai is the unit of tone placement, but it can best be reached by way of the "syllabic base" minus prosodic features, as proposed by Eli Fischer-Jørgensen: "In most languages this syllabic base may be defined structurally as the class of the smallest units, of which each (in connection with stress, tone and intonation, if such units are distinctive in the language in question) is capable of constituting an utterance by itself."[3] This does not imply, she goes on to say, that all members of the class actually occur as utterances. It implies that some are not found because of accidental gaps in the inventory of signs. The phonemic syllable in Thai, then, is this syllabic base comprised of inherent features combined with a simultaneous tonal feature. The simultaneity may be either partial, with the tonal feature combined with just the vocalic nucleus, or total, with the tonal feature spreading into whatever voiced consonants are present in the base.

In this connection, it is useful to recognize the three styles of speech — isolative, combinative, and rapid combinative — posited for Thai by Eugénie J. A. Henderson.[4] The phonemic syllable in the isolative style, which is very slow and careful but not over-differentiated, is taken as the frame of reference for the phonemes and much of the acoustical work. Other parts of the phonology, of course, as well as such matters as tonal sandhi, cannot be presented without reference to the combinative styles.

2. Systematic Inventory of Phonemes

The phonemes are presented in subsystems in terms of inherent vocalic and consonantal features and tonal features. This is followed by an amalgamation in the form of a systematic display.

a. Vowels

	Front	Central	Back
High	i	ɨ	u
Mid	e	ə	o
Low	æ	a	ɔ

Vowel length

Vocalic duration is phonemically distinctive. Long vowels are here interpreted as sequences of two vowels because of a distributional parallel with the three vocalic clusters /ia ɨa ua/. Short vowels, geminate vowels, and vocalic clusters all take unitary tones. Examples are /mia/ 'wife', /hǔa/ 'head', /mɨ̂a/ 'when', /hàt/ 'to practice', /hàat/ 'shore'. This interpretation agrees with that of Mary R. Haas,[5] Josephine A. Gillette,[6] Mary E. Kroll,[7] and William J. Gedney[8] but differs from that of other scholars. A. Haudricourt[9] posits "short" and "long" vowels and appears to consider the vocalic clusters monophonematic long vowels. R. B. Noss[10] posits a phoneme of "weak voicing" for short vowels, and Foongfuang Kruatrachue,[11] a phoneme of "reduction of length." G. L. Trager[12] is unique in considering long vowels sequences of V + /j w/ or /h/. All these solutions can be justified logically, of course, but the measurements and experimental results of Chapter II give strong support only to those that regard relative vowel length as distinctive.

Allophones[13]

The major phonetic interest in vowels in this study is an acoustic one; however, it may be of interest to mention some articulatory variation at this point.

There is some qualitative variation accompanying quantitative distinctions. /i/ is usually [ɪ] except in final position where it is [i], but /ii/ is [i:]. /e/ is [ɛ] with occasional occurrences of [e], but /ee/ is [e:] varying to [ɛ˔:]. /a/ has an allophonic range including [ʌ˕ ɐ ɑ˔ ɑ+], [ɑ+] being most typical of the isolative style, while /aa/ is [ɑ+:]. /u/ is usually [ʊ] except in final position where it is [u]; /uu/ is [u:]. /o/ ranges from a somewhat unrounded [o˕ᶜ] to [o]; /oo/ is [o:].

No consistent qualitative variation appears to accompany the quantitative contrast in the other vowels. /æ/ is [æ˖]. /ɨ/ is a high central unrounded [ɨ] but also occurs as a high back unrounded [ɯ].[14] /ə/ ranges from upper mid-central unrounded [ɜ] to [ɤ]. /ɔ/ is [ɔ].

Vowels after nasal consonants and /h/ are often slightly nasalized.

b. Consonants

Occlusives

Stops

	Voiceless		Voiced
	Plain	Aspirated	
Labial	p	pʻ	b
Dental	t	tʻ	d
Velar	k	kʻ	–

For some speakers /kʻ/ has a velar affricate allophone [kˣ].

There is a "hole in the pattern" in the voiced velar spot; however, Haas posits a /g/ occurring only finally. Since there is no opposition between plain, aspirated and voiced stops of the same localization in final position, the view taken here is that there is a neutralization of the manner features at the end of a syllable with the archiphonemes, written as /p t k/, occurring as [p t k] or [b d g]. Pre vocalic /p t k/ are pharyngealized and tend to lower and centralize following vowels somewhat. Trager[15] reduces the inventory by positing sequences of voiceless stop + /h/ for the aspirated stops and voiceless stop + /ʔ/ for the voiced stops.

Haas, as well as Fowler,[16] puts a /ʔ/ among the stops. The status of the glottal stop is problematic. It is here considered a feature of internal open juncture /+/ occurring before V or VV and after V. It is in free variation with zero in initial prevocalic position and at any terminal juncture after V, VV̂, and VV́. The very few special cases of onomatopoeia and Chinese loans in which morphemes have [ʔ] after a vocalic cluster, e.g. [pʰìɐʔ] 'slapping sound', constitute a marginal phenomenon.

The phonemic status of [ʔ] can be supported on the basis of pattern congruity.[17] Where final [ʔ] occurs, there are the same limitations on the privileges of occurrence of the tones as in the presence of final /p t k/. With /ʔ/ in the inventory, there would be symmetrical tonal behavior in such pairs as

/sàk/ 'only' vs. /sàʔ/ 'to shampoo'
/lák/ 'to steal' vs. /láʔ/ 'to leave'.

Appealing as this argument is, the unstable nature of [ʔ] makes it difficult to accept as a distinctive phoneme. Indeed, informants accept such renditions as [ʔà:n] and [à:n] 'to read' and [pʰráʔ] and [pʰrá] 'priest' as "the same" and "equally good" pronunciations.[18]

Trager's positing of /ʔ/[19] is, it would seem, an example of the manipulation of symbols to reduce the inventory. /ʔ/, in his view, is not just manifested as the phone [ʔ] but also as the voicing of final stops before initial vowels of following syllables. This leads to the elimination of /b d/ as well as Haas's /-g/, leaving the clusters /pʔ tʔ kʔ/ for [b d g]. It is felt that this approach, while internally consistent and, at least on one level, economical, is too great a departure from phonetic reality to be followed here.[20]

Affricates

	Voiceless		Voiced
	Plain	Aspirated	
Palatal	c	cʻ	–

The palatal affricates are wrongly described by Fowler as voiced and voiceless respectively.[21] At the intersection of the features of voicing and palatality there is a "hole in the pattern." This hole is filled by an occasional variant of /j/. [ʃ] is an occasional variant of /cʻ/.

/c cʻ/ are interpreted as the clusters /tj thj/ by both Noss and Trager.

Together, the stops and affricates make up the class of occlusives.

Laryngal	h

/h/ can be regarded as the glottal correspondent to the feature of aspiration and aligned with the aspirated stops. Pointing out that if the cluster interpretation of the aspirated stops is rejected, /h/ is apparently a manner consonant in Thai, Hockett says, "If [pʻ] is a unit phoneme /pʻ/. . . . , and there is an /h/ in the system, then (other things being suitable) we say that /h/ is a manner consonant, going with the aspirated stops."[22]

Nasals

Labial	m
Dental	n
Velar	ŋ

The nasals have the same localization features as the stops.

Fricatives

Labial	f
Dental	s

There are only voiceless fricatives, but an occasional variant of /w/ provides /f/ with a voiced partner [v] or [β].

Liquids

	Lateral	Flapped
Dental	l	r

The liquids are localized in a general way in the dental region.

The phonemic status of the [l] and [r] phones is problematic, although other writers, except Gedney, make little of it.[23] For Thai observed in the normal daily intercourse of the general population of Bangkok as well as among speakers of this dialect in other provinces of Thailand, the following seems to be a fair statement. One could say that there is a phoneme /l/ with an allophonic range that is rather wide to the Western ear. It may be realized as any of a number of phonetically related sounds ranging from a lateral [l] at one extreme through a continuum of variants to a trilled apical [r] at the other extreme. Two intermediate points are retroflex [r] and flap [ɾ]. [l] is the most favored allophone in all environments except after /t/, where the most likely articulation is [ɾ] or [r].

Since the Thai writing system provides separate letters for l and r, many of the highly literate claim that they do indeed maintain an /l/: /r/ opposition at least when they speak "correctly". Such people, however, are often to be heard asking each other whether a word is spelled with l or r. Even the question has its pitfalls. In asking whether a word is spelled with l as in /liŋ/ 'monkey' or r as in /rɨa/ 'boat',[24] the speaker is likely to pronounce [l] in the names of both letters (/lɔɔ/ and /rɔɔ/) and in the two key words! As members of this group, school teachers are constantly reminding their pupils to differentiate /l/ and /r/, and in classroom dictations they tend to use a heavily trilled [r] to announce the presence of r in the spelling. It is true, however, that some members of this literate group, particularly those in the upper classes,[25] more or less consistently maintain an /l/:/r/ opposition in formal literary speech and even in daily conversation.

The following excursus is an experimental investigation of the distinction among highly literate Thai speakers, as typified by the informants who cooperated in this study. It is included here, because it has some bearing on the positing of the phonemic distinction. One identification test and five paired-utterance tests were prepared on magnetic tape.

Test A. — The morpheme /lót/ 'to reduce' was recorded by an informant in two unambiguous contexts:

1. cà lót raakʻ aa hâj tʻ ân ↓
   'I'll reduce the price for you.'
2. lót raakʻ aa mâj dâj ↓
   'I can't reduce the price.'

This was also done with /rót/ 'vehicle':

1. paj rót kʻ an diaw ↓
   'We'll go in a single car.'
2. rót paj mâj dâj ↓
   'The car won't start.'

Each of the two morphemes was cut from its environments[26] and copied ten times, making a total of forty stimuli. These were arranged in random order and presented to seven Thai subjects for identification in terms of two definitions written in Thai: (1) /dɨŋ loŋmaa/ 'to reduce' and (2) /pʻ aahàná/ 'vehicle'.

Test B. — The morphemes /lôok/ 'world' and /rôok/ 'disease' were each taken from two unambiguous contexts. Each of the four morphs was combined with each of the other three and with itself in two orders, making twenty pairs. Two copies of each pair, making forty test stimuli, were arranged in random order and presented to six subjects, who were told to judge the members of each pair as the same word or different words.

Test C. — /lót/ 'to reduce' and /rót/ 'vehicle' were each taken from one unambiguous context and combined into four pairs: /l l/, /l r/, /r l/, /r r/. Five repetitions of each pair, making twenty test stimuli, were presented in random order to six subjects as a paired-utterance test.

In Tests D-F minimal pairs were formed from a scrambled list of isolated words written in Thai script with initial *l* and *r*. The informant recorded the words one by one. They were provided with glosses.

Test D. — /laán/ 'to be bald' and /raán/ 'shop'. 4 pairs X 5 = 20 stimuli (paired utterances) in random order. 5 subjects.

Test E. — /læ̂æk/ 'to swap' and /ræ̂æk/ 'to be first'. As in Test D. 4 subjects.

Test F. — /luaŋ/ ‘deceive’ and /ruaŋ/ ‘ear of paddy’. As in Test D. 4 subjects.

In all the tests only those recordings were used that contained clear [l] phones for the posited phoneme /l/ and clear [r] phones for the posited phoneme /r/.

The test results are given simply as percentages of “correct” responses, since internal details can have no bearing on the general conclusion.

Tests:	A	B	C	D	E	F	Average
Subjects tʿ.cʿ.	57.5	–	–	–	–	–	57.5
k. s.	97.5	90.0	90.0	75.0	60.0	100	85.4
u. s.	30.0	82.5	70.0	–	–	–	60.8
s. r.	75.0	90.0	100	90.0	–	–	88.8
e. n.	100	80.0	95.0	85.0	65.0	100	87.5
c. s.	97.5	77.5	90.0	85.0	60.0	85.0	82.5
b. w.	80.0	80.0	85.0	90.0	95.0	80.0	85.0
Average	76.8	83.3	88.3	85.0	70.0	91.0	78.2

These results certainly do not justify the elimination of /r/ from the phonemic inventory of Standard Thai, whatever the situation may be for the bulk of illiterate and semi-literate speakers of the central dialect. On the other hand, an average of only about 80% for the whole table suggests that the /l/ : /r/ opposition is not too stable. In addition, it must be borne in mind that the test design favored the showing of a distinction, since much of the range of phonic variation actually occurring in normal speech, and indeed in the recordings, was excluded. Had a morpheme like /raân/ ‘shop’ been allowed to be represented by [lá:n] and contrasted with [lá:n] representing /laân/ ‘to be bald’, there would not have been much point in giving the test.

Taking all the foregoing social and linguistic facts into account, one would probably be safest in accepting the presence of /l/ and /r/ in the system, as posited at the head of this section, while noting the unstable nature of this opposition and the probability of its being in a state of flux.[27]

Semivowels

Labial	w
Palatal	j

The chart of phonemes is arranged to show how the semivowels can be linked to the vocalic system. /w/ is normally [u̯] but has occasional variants in [v] or [β]; /j/ is normally [i̯] but has occasional variants in the voiced palatal affricate [dʑ] (See Affricates and Fricatives above).

c. Tones

Dynamic		Static	
		High	/
Falling	∧		
		Mid	0
Rising	∨		
		Low	\

In the dynamic tones the upward and downward movements are relevant, not the end points. The static tones are analyzed as three levels. /O/ is not marked in the transcription. Except for the internal organization, this accords with the usual view[28] that there are five phonemic tones in the language. Noss, however, apparently because he considers the emphatic high tone phonemic, sets up six tones. The emphatic tone is treated here as a non-distinctive emphatic feature.

Kroll also sets up six tonal phonemes, but her sixth one is "lack of tone" occurring only in plurisyllabic utterances. She very rightly gives special status to these "toneless" syllables,[29] but the matter can be handled without adding to the inventory of phonemes. The syllables involved are of the type ending in a single vowel, which takes either /ˊ/ or /ˋ/ in the isolative style. In the combinative style of speech the /ˊ/:/ˋ/ opposition is neutralized, and the pitch of such a syllable varies about the norm for the mid tone, moving above or below it under the influence of surrounding tones. This neutral tone is conveniently treated as a realization of the mid tone and written accordingly. In a morphophonemic writing one would indicate which tone, /ˊ/ or /ˋ/, occurs on the syllable in isolative speech. Minimal pairs showing this neutralization are hard to find, but it can be demonstrated in analogous environments. /à/ in /k‘àmeěn/ ‘Cambodian’ is opposed to /á/ in /k‘ámeen/ (as in /hòkk‘ámeen/ ‘to tumble’) in isolative speech. In the combinative style /k‘a/ appears in both. In her *Thai Vocabulary* (Washington, D. C., 1955) Haas lists such words with their morphophonemic alternants.

Stating that this neutral tone is a separate distinctive phoneme obscures the structure of the language. It is nothing but the neutralization of static tonal features occurring for the most part on open single-vowel syllables of plurisyllabic morphemes in combinative speech; it occurs also on some monosyllabic morphemes like /cà/ ‘will, shall’.

In line with his general economy of inventory, Trager reduces the tones to three — high, mid, and low — which occur singly and in sequences of two. He suggests that this scheme "not only opens the possibility of accounting for reported dialect differences in one overall system, but also simplifies the statements of morphophonemic tone adjustments."[30] As for the overall system, it is felt here that more descriptions of the many dialects are needed before much can be said about it.[31] In spite of the contention of morphophonemic simplicity, it seems rather that the phonetic variation of the pitch contours in running speech would make the assigning of the end points to any of three tones much too difficult no matter how rigorously logical it is on paper.

Allophones

The phonetic nature of the tones is a major topic of the acoustical study; therefore, only a few points will be made here.

In final position[32] the falling tone starts above the middle of the voice range, rises slightly and then drops abruptly to the bottom of the range; the rising tone starts below the middle, drops slightly, and then rises abruptly to the top. In medial position the extent of the fall and rise of the dynamic tones varies somewhat under the influence of the speed of utterance and the following tone as well as the general sentence intonation.

In final position the high tone starts fairly high in the voice range, rises slightly and then drops a little at the end with concomitant glottal constriction setting in before the end, except in single-vowel syllables ending in Ø or a stop, where the final drop and the constriction are absent. The mid tone starts around the middle of the voice range and remains level until it drops slightly at the end. The low tone starts fairly low and drops slightly. In medial position the terminal drops of the upper two static tones and the constriction of /´/ are absent.

Morphophonemic alternation, practically all of the automatic kind, will not receive separate treatment here.

d. Table of phonemes

All the phonemes are presented together in terms of distinctive features in the table. The vowel quadrilateral is turned on its side to show the relation between the labial consonants and the back vowels, and the palatal consonants and the front vowels. The dashes in two of the boxes mean that there is a variant of /j/ at the intersection of the voiced occlusive and palatal features and a variant of /k/ at the intersection of the voiced occlusive and velar features.

3. Distribution of Phonemes

Although it makes no difference for the consonants, dividing syllables into single-vowel and double-vowel types helps in stating the distribution of the vowels and tones.[33]

a. Syllable types

1. $(C_1(C))\overset{T}{V}(C_2)$
2. $(C_1(C))\overset{T}{V_1V_2}(C_2)$

b. Consonants

C_1 = any consonant

C = { w / l / r } in the clusters: { k‘w kw / k‘l kl p‘l pl / k‘r kr p‘r pr t‘r (rare) tr }

C_2 = nasals, stops, semivowels,
but /w/ only after single or geminate / i e æ ə a/ and /ia/, and /j/ only after single or geminate /ə a u o ɔ/ and /ɨa ua/.

c. Vowels

Syllable Type 1.

V = any vowel except for limitations imposed by final /w j/.[34]

Syllable Type 2.

V_1 = any vowel except when V_2 is different.

V_2 = a. V_1 doubled.
b. /a/ as the second element in the clusters /ia ɨa ua/.

THE PHONEMES OF THAI

TIMBRES

	CONSONANTS								VOWELS			
	Liquids		Fricatives	Nasals	Occlusives			Semivowels	High	Mid	Low	
					Voiceless		Voiced					
	Lateral	Flapped			Asp.	Plain						
Labial			f	m	pʻ	p	b	w	u	o	ɔ	Back
Dental	l	r	s	n	tʻ	t	d	■	ɨ	ə	a	Central
Palatal					cʻ	c	←	j	i	e	æ	Front
Velar				ŋ	kʻ	k	→					
Laryngal					h							

TONES

	Dynamic	Static	
Falling	^	´	High
		o	Mid
Rising	ˇ	`	Low

■ Non-alignment of dentals and central vowels

← Overlap of free variants of /j/

→ Overlap of free variants of /k/

d. Tones

Syllable Type 1.
T = a. /´/ or /`/ if C_2 is a stop or ϕ. (/kʻâ/ ʻfemale polite particle' is among the rare exceptions.)[34]
b. all 5 tones with all other members of C_2.

Syllable Type 2.
T = a. /^/ or /`/ if C_2 is a stop. (/noót/ ʻnote', an English loanword, is among the rare exceptions.)
b. All 5 tones with all other members of C_2, including ϕ.

4. Syntagmatic Phonology

a. Internal open juncture

$\boxed{+}$

When an intervocalic consonant or permitted consonant cluster is in normal transition at a syllable boundary, it is the initial element of the following syllable. Examples are /pràtuu/ ʻdoor', /pʻuûjàj/ ʻadult', /tàkraj/ ʻscissors'. Otherwise, certain demarcative features serve to assign such an intervocalic consonant or the first consonant of a permitted cluster, or a vowel in a sequence of vowels, to the preceding syllable. These features are collectively labelled plus-juncture, a unit of the syntagmatic phonology rather than a distinctive phoneme. Plus-juncture is realized as the presence of a final segmental allophone before it and an initial one after it or by the presence of [ʔ]. Examples: [sɐ̀ʔ ɑ̀:t] /sà+aàt/ ʻto be clean', [kâᵘ̯ʔî:] /kâw+iî/ ʻchair', [kɑ:nʔɑ̀:n] /kaan+aàn/ ʻreading', [nɑ́k˺rɑm] /nák+ram/ ʻdancer'.

Since plus-juncture is here taken to be a syntagmatic phonological feature, its occurrence will be indicated in the transcription only where syntagmatic cuts are not otherwise apparent. With word boundaries, whether or not phonologically manifested, already shown by spaces in the transcription,[35] plus-juncture will be written only within plurisyllabic words. A sequence of two consonants forming an unpermitted cluster will unambiguously mean that the syllable boundary falls between them; e.g. /kʻonŋaan/ ʻworker'. For the most part /+/ occurs at morpheme boundaries.

It is to be noted that in other treatments, for example Kroll's where word boundaries are not recognized, plus-juncture is indicated at many points in the transcription where its phonetic justification is rather tenuous.[36] Haas, who has word boundaries in her transcription, as well as /ʔ/ and final /b d g/ in her inventory, is not confronted with problems of internal open juncture.

b. Sentence intonation

Pitch Registers		Terminal Junctures	
High	$\boxed{2}$	Rising pitch	$\boxed{\uparrow}$
Normal	$\boxed{1}$	Sustained pitch	$\boxed{\vert}$
		Falling pitch	$\boxed{\downarrow}$

The interesting subject of intonation in a tone language has been discussed by a few investigators,[37] but the matter of analyzing it into discrete units in anything like a thoroughgoing way has apparently remained problematic.[38] In two recent studies of Chinese dialects, intonation receives considerable attention. Søren Egerod's statement for Lungtu[39] is quite clear and generally applicable to Thai: "The sentence intonation is superimposed on the individual tones and determines the absolute pitch and general inflexion of the utterance, whereas the tones determine the relative difference in pitch and inflexion among the syllables. The sentence intonation influences and modifies the pitch and inflexion of the tones. Only at the end of an utterance may the tone be modified beyond recognition by the intonation." He then says that Lungtu utterances tend to end on middle even or high even pitches, making the last syllable identical with the middle or high tone. He accordingly marks intonation simply by showing the tonal adjustment of the final syllable. Going into a detailed account of the nature and uses of intonation in Chengtu, Nien-Chuang T. Chang ends up with two patterns of perturbation of the tones of final syllables: a "rising tune" and a "falling tune".[40]

Although Thai intonation has been treated at considerable length by Kroll and Gillette, their phonemic treatments are not too convincing. When actually functioning in daily communication, Thai intonation is seen to be an essentially continuous, ungraded phenomenon, hardly susceptible of analysis into distinctive phonemes.[41] Under the influence of the attitude and condition of the speaker, the pitch level varies considerably through the course of a conversation. The distinctive tones can be regarded as levels and movements superimposed on an underlying varying voice register. No attempt at a thorough description of this elaborate interplay of tones and intonation contours is made here. Indeed, intonation as such is not included in the acoustical investigation, although it would be a logical continuation of the present study. It is hoped, rather, that the simple scheme presented here will indicate an awareness of the problem and help fill out the operational framework, which is the goal of Section B.

A Thai intonation contour consists of one of two voice registers and one of three terminal junctures. The three terminal junctures occur at the ends of sentences and at other syntactic breaks. The voice is either on a low register /^{1}/, the normal pitch range of speech, or a high register /^{2}/. Only the high register is marked in the transcription. At the juncture of falling pitch /↓/ the pitch drops slightly and the voice dies away. E.g.,

pen kʻon dii↓ ʻHe's a good person.'
pen kʻon dii↓ tæ̀æ cʻán mâj cʻɔ̂ɔp ↓ ʻHe's a good person,
but I don't like him.'

At the juncture of sustained pitch /|/ there is a pause or drawl with sustension of the preceding register. E.g.,

pʻuût pʻasaǎ tʻaj | dâj dii↓ ʻHe can speak Thai well.'

At the juncture of rising pitch /↑/ there is a rise in pitch which signals surprise, doubt or a question. E.g.,

dii↑ ʻIs it good?'
kʻǎaw↑ ʻIs it white?'
^{2}dii↑ ^{2}mâj cʻɨ̂a↓ ʻGood? I don't believe it!'[42]

The presence of /↑/ may so affect the lexical tone on the final syllable as to make it indistinguishable from /ˇ/.

It is now convenient to mention that the frame for the general statement on tonal allophones was /^{1} ____↓/. Before /|/, /O/ and /ˊ/ do not have a final drop, and /ˋ/ and /ˆ/ drop less than before /↓/. Before /↑/, /ˆ/ does not have its initial rise and /ˇ/ its initial fall; the static tones all rise slightly at the end; /ˊ/ does not have final glottal constriction. In the presence of /^{2}/ the absolute pitch of the tones is higher. Although the matter has not been fully investigated, it appears that certain tonal oppositions may be neutralized at terminal junctures.

Two linguists who have described Thai intonation briefly in non-phonemic terms are Haas and Henderson. Haas speaks of the slight lowering of the voice at the end of a phrase or sentence[43] and gives examples of allophonic differences in tones in this position. In her discussion of sentence prosodies, Henderson[44] says much the same thing on intonation but gives a detailed account of "sentence tone", which is the selection of modal and polite particles for certain syntactic breaks, especially the ends of phrases and sentences. "Particles bearing sentence tone are distinguished from other monosyllables in the sentence in one important respect: the disposition of tone and quantity is determined not by the phonetic structure of the particle itself, but by the requirements of the sentence as a whole." This analysis takes care of most of the exceptions to the tonal distributional rule for syllables with single vowels (3. Distribution of Phonemes). Since the frequent nonoccurrence of final particles makes this explanation insufficiently general, it is felt that a Thai grammar would do well to use Henderson's penetrating description of particle selection in conjunction with some such broad scheme for handling sentence intonation as is advanced here.

5. Emphatic Features

Much in the way that the term <u>Emphatika</u> was used by Laziczius,[45] non-distinctive phonological features that give prominence to a portion of the speech chain are considered emphatic features. Much remains to be done in this area of phonology. Here only two such features in Thai, emphatic tone and contrastive stress, will be considered.

a. <u>Emphatic tone</u>

$\boxed{˝}$

In verbal reduplications, the emphatic tone may appear on the first morph of the construction, replacing the normal tone and indicating intensification of meaning. It also occurs in exclamations and is favored by women. The emphatic tone has a high pitch, higher than / ́/, and is accompanied by lengthening of the syllable. Examples:

di̋idii↓	'How good it is!'
rɔ̋ɔnrɔ́ɔn↓	'How hot it is!'
wa̋anwǎan↓	'How sweet it is!'

M. Haas, who treats this matter in some detail,[46] considers / ̋/ a modification of the high tone.

b. <u>Contrastive stress</u>

$\boxed{|}$

Contrastive stress makes any syllable in the speech chain prominent through extra loudness, a rise in pitch or lengthening of the vowel, or some combination of these features. E. g.,

^{2}pen ˈkʻruu↓ 'He's a teacher.'

Noss and Kroll posit three stress phonemes. Indeed, Kroll handles sentence intonation by means of volume contours and terminal junctures. It is difficult to see, however, why any distinctive value should be ascribed to stress in Thai. Manifestations of stress other than /ˈ/ appear to be predictable, since a syllable on the "neutral tone" (2c. Tones) automatically takes a weaker stress than surrounding syllables. The remarks on intonation as a continuous, ungraded phenomenon are also applicable to attempts at setting up levels of stress in Thai.

In addition to the foregoing phonological features, certain grammatical devices are also used to express emphasis, but these, of course, are different phenomena.

C. Experimental Techniques and Instrumentation

An account of the instruments, techniques and personnel used in this study is presented in one place for the convenience of the reader.

1. Instruments and Their Use

Each instrument is described briefly without highly technical details; more information can be found in the sources cited.

Tape recorders

Magnecorder PT6A and Rangertone "professional" recorders were used at a speed of 15 or 7.5 inches per second for playing speech into analyzers and preparing tests. A Revere T1100 portable recorder was used at a speed of 7.5 i.p.s. for the recording of speech and the administering of tests away from the laboratory. This machine was judged to have more than enough fidelity for speech work.

Tape cutting and splicing techniques were used to isolate stretches of speech from the rest of an utterance. To locate an end point it is often enough just to run the tape back and forth by hand over a magnetic head while listening to the sound. Watching the changing level on a VU meter or the changing wave shape on an oscilloscope is helpful. This procedure was used to find the end points of vowels so that measured segments could be cut out for the experiments on vowel duration (Chap. II). Since no cuts were made right at the points marked, this kind of cutting and splicing was accurate enough.

For more precise localization magnetic clicks were used. A magnetized knifeblade is drawn across the tape a few inches from either end of the stretch to be segmented, and the tape is marked at these points. Two such points serve as a check on each other. This results in a pair of clicks which appear as vertical black lines on a spectrogram. The points at which cuts are to be made are marked on the spectrogram. The distances from each mark to the two clicks are measured and then laid out on the tape by means of a mathematical conversion: 1 second = 15" on the tape and 5" on the spectrogram. Perpendicular cuts ("butt-splices"), which will not touch any of the environing speech, are made to remove the segment from the tape and splice it onto another. As a safety measure, it is best to make several copies of the original recording before starting this procedure. Magnetic clicks were used for isolating words in /r/ and /l/ from sentences (Sec. B. 2.b).

Another kind of tape playback device used was the Ellamac Language Master,[47] which plays back a strip of magnetic tape that has been cemented to a card. This "card reader" makes it convenient to listen repeatedly to short utterances and make ready comparisons of them. The manufacturer's non-standard speed was altered to 7.5 i.p.s. so that cards could be prepared easily in the laboratory. The card reader was used for the experiments on the perception of vowels (Chap. I).

Sound spectrograph[48]

The sound spectrograph used for this study was the Kay Sonagraph[49] with certain modifications. Speech of no more than 2.4 seconds in duration is recorded on a magnetic drum within the spectrograph. The speech is played back repeatedly at high speed while a pass-band filter slowly moves across the frequency spectrum. The filter can be set at a "wide" band of 300 cycles per second or a "narrow" band of 45 cps. The filter output is connected to a stylus that traces a frequency-intensity-time pattern on a sheet of electrically sensitive paper wrapped around a revolving drum. Frequency is shown vertically and time, horizontally. Gradations of blackness in the spectrogram give an approximation of relative intensity, and a section showing amplitude versus frequency can be made for any point in time. One can also make an overall amplitude display along the horizontal axis above the spectrogram.

The standard Sonagraph is equipped with a microphone input and a frequency scale of 2000 cycles per inch. The spectrograph used in this study was modified to take speech directly from a tape recorder and to provide two more scales of 200 and 1200 c/in.

Wide and narrow band spectrograms at 1200 c/in., supplemented on occasion by sectional views, were used for the determining of the formant frequencies of the vowels (Chap. I) and the measuring of vowel durations (Chap. II). Narrow band spectrograms at 200 c/in. were used for finding the tonal contours (Chap. III). The segmentation of utterances for the work on /r l/ was done with the aid of spectrograms (Sec. B-2b). Amplitude displays tracing the course of intensity changes were used for supplementary information on tonal features.

Because of slight fluctuations in the behavior of the spectrograph, not unusual in electronic apparatus, the scales could not be trusted to remain permanently at their nominal values. A rather precise calibration of the 1200 c/in. scale was made for each group of formant measurements by making a spectrogram of a complex wave of 400 cps and measuring the harmonic spaces. A bit of this calibrating tone was recorded in the spectrogram with each speech specimen to furnish reference points from which formant frequencies could be found.

The same thing was done for the 200 c/in. scale, except that a complex wave of 100 cps was used so as to have a number of harmonics in the small portion of the spectrum displayed. This was especially important because this scale was non-linear, and a calibration curve based on the gradually diminishing harmonic spaces had to be used.

Pattern Playback[50]

The Pattern Playback is a speech synthesizer that uses an optical system to convert spectrographic patterns to sound. An intense thin line of light is focussed radially on a rotating "tone wheel", a film transparency bearing fifty concentric sound tracks. The disc is rotated at such a speed that the innermost track responds as a tone of 120 cps and the others as multiples thereof up to 6000 cps. The modulated line of light from the tone wheel is focussed on a moving sheet of clear acetate bearing a hand-painted spectrogram. The vertical axis of the spectrogram is divided into fifty channels, each of which represents a multiple of 120 cps; thus the first channel is 120 cps, the second, 240 cps, and so on. Light which has been modulated at the frequencies corresponding to the channels in which paint is present is reflected into a common phototube. The photocurrent is then amplified and fed into a loudspeaker or a tape recorder.[51]

Thus the timbre and dynamics of the sound are controlled by the placement of paint along the frequency scale of 1200 c/in. and the time scale of 7.2 in./sec. Widening a band of paint brings in more channels and increases the intensity of the sound. Tapering the ends of a band produces a gradual change in intensity. Since the fundamental frequency is fixed at 120 cps, the pitch of the output does not vary.

The Pattern Playback was used to synthesize vowels (Chap. I).

Vocoder[52]

The 18-channel Vocoder, a machine that was originally developed for use in telephonic communication, processes speech, first by analyzing it and then by synthesizing it. The analyzer separates the speech signal into information about its voiced and voiceless characteristics and its spectrum, or roughly, into information about activity at the vocal cords versus information about articulation. The synthesizer supplies its own voiced sounds in the form of a buzz, which can be set to follow the pitch changes of the original voice or remain a monotone, and its own voiceless sounds in the form of a hiss. These sounds are shaped by the spectrum information from the analyzer. The result is speech that is highly intelligible but has a voice quality that is characteristic of the machine.

The Vocoder was used briefly for the experiments on vowel perception (Chap. I).

Intonator[53]

The Intonator is a device that permits the manipulation of pitch as speech passes through the Vocoder. Three loops, a magnetic tape bearing the speech, a pitch-control tape and a reference spectrogram (optional), are kept in step by means of sprocket holes. A pitch contour painted on the pitch-control tape varies the fundamental frequency (i.e., the voice pitch) of the speech synthesized by the Vocoder; all other characteristics conform to those of the original speech recording. One can, by these means, impose a pitch pattern on an actual utterance in a way that was never intended by the original speaker.

The Intonator was used to synthesize tones in the experiments on the perception of tonal features (Chap. III).

2. Subjects

The Thai personnel cooperating in this study served both as linguistic informants and as subjects in psycho-acoustic tests.

They were all professional people, mostly school teachers, studying at local universities.

The subjects were native speakers of Standard Thai. Some had been born and raised in Bangkok. Others had been brought to the capital in early childhood and brought up there. A few spent their childhood and early youth in the provinces where they grew up as bidialectals, switching between the local dialect and the standard as occasion demanded. They had at least the latter part of their secondary education and all their university education or professional training in Bangkok.

Informants

The three principal informants for the linguistic analysis given in Section B were Mr. Ekawit Nathalang (E. N.), Miss Khomkhaay Chongcharoensuk (K. C.) and Miss Kandaa Sindhvananda (K. S.). Several others had been used previously in Thailand.

Mr. Ekawit Nathalang, together with his brother Mr. Wathanyuu Nathalang (W. N.), supplied most of the voice recordings for acoustical analysis. Samples were also taken from several others.

Psychological experimental group

Fourteen men and eleven women served as subjects for the psycho-acoustical tests. A gradual turnover in personnel was caused by the departure of some subjects and the arrival of others during the course of the study.

3. General Testing Procedure

The stimuli for each perception test were arranged in random order by drawing numbered poker chips from a bag. This resulted in a tape recording with typical time intervals of four seconds between stimuli and ten seconds after every ten stimuli. In most of the tests the subjects were told to identify each stimulus as a Thai expression, which they wrote on their answer sheets in Thai orthography. Variations on this general procedure are noted where relevant.

Schedule conflicts usually made it necessary to test only one or two subjects at a time and seldom more than three; consequently, steps were taken to insure uniformity of instructions for all presentations of a test, and the same recording was used each time.

Notes

1. Informants and test subjects were either native speakers of the dialect or bi-dialectals who had grown up speaking both Standard Thai and some local dialect. They are described at greater length in Section C-2 of the Introduction.

2. Noss, Gillette, Kroll, and Trager. See below.

3. "On the Definition of Phoneme Categories on a Distributional Basis," *Acta Linguistica*, VII (1952), 16. This is not the place to enter into a discussion on the notion of the syllable. This definition seems to work for Thai.

4. "Prosodies in Siamese: A Study in Synthesis," *Asia Major*, N.S. I (1949), 189.

5. "A Brief Description of Thai with Sample Texts," *Outline for Types of Linguistic Structure* (Berkeley, 1956), p. 2.

6. "Prosodic Features in Bangkok Thai" (unpub. master's thesis, Georgetown U., 1955), p. 1.

7. "Suprasegmental Phonemes of Thai (Bangkok Dialect)" (unpub. master's thesis, Georgetown U., 1956), pp. 2-3, but she subsumes the second element of vocalic clusters under a "centering glide" /h/.

8. "Indic Loanwords in Spoken Thai" (unpub. Ph. D. dissertation, Yale U., 1947), pp. 11-13, but he posits two degrees of length for the vocalic clusters as well.

9. "Les phonèmes et le vocabulaire du thai commun," *Journal asiatique*, CCXXXVI (1948), 200-201.

10. "An Outline of Siamese Grammar" (unpub. Ph. D. dissertation, Yale U., 1954), Sec. 1.2.1.

11. "Thai and English: a Comparative Study of Phonology for Pedagogical Applications" (Ed. D. dissertation, Indiana U., 1960), p. 33.

12. "Siamese Phonemes: A Restatement" in *Studies Presented to Yuen Ren Chao* (Taipai, 1957), p. 25.

13. IPA symbols are used for allophones.

14. E. Henderson labels it [ɯ] without comment in "Specimen: Thai," *Le maître phonétique*, LXIX (1940), 11-12.

15. Trager, pp. 22-23.

16. M. Fowler and Tasniya Israsena, *The Total Distribution of the Sounds of Siamese* (Madison, 1952), p. 4.

17. P. K. Benedict, "Tonal Systems in Southeast Asia," *JAOS*, LXVIII (1948), 186.

18. See Gedney, p. 26, for a detailed statement on the glottal stop that arrives at essentially the same position. In a recent letter, however, Gedney indicates that neither he nor I nor anyone else has fully accounted for the facts of the Thai glottal stop.

19. Trager, p. 23.

20. Trager's unusual use of the term "glottal stop" makes his reasoning difficult to interpret. He cites authors who use it in its conventional phonetic acceptance for Thai but then describes it as "the initial onset of voicing." In general, it is hard to see how the stops [p t k] could assimilate any voicing from an abrupt, tight closure of the vocal folds. If, on the other hand, the folds are actually vibrating for the production of voice, it would seem to be a real strain on the notion of phonetic reality to classify this with occurrences of genuine glottal stop. In actual fact, in an utterance like /sàbaaj/ 'to be well', spoken normally without internal open juncture, voicing begins in the vowel before the [b] phone and is not interrupted. Where, then, is the glottal stop?

21. Fowler, p. 4.

22. Charles Hockett, *Manual of Phonology* (Baltimore, 1955), p. 125.

23. In M. R. Haas and H. R. Subhanka, *Spoken Thai* (New York, 1945), p. 55, students are cautioned that a given informant may interchange /l/ and /r/ and should be imitated even when he departs from the transcription. Kroll, p. 17, says that her informants were speakers of standard Thai, sharing "the characteristic of interchanging the phonemes /l/ and /r/." In her phonetic transcriptions, Gillette writes /r/ as [l] apparently wherever her informant said it that way but makes no comment on the matter. Gedney, p. 15, shows that /l/ : /r/ is not a stable contrast. See also Kruatrachue, p. 47.

24. School children learn to name letters together with traditional key words that resolve ambiguity where two or more letters represent the same phoneme. Where no such ambiguity exists, people will often just name the letters without bothering with the key words.

25. No sociological definition can be attempted here. See Gedney, pp. 15-16, for another discussion of the sociolinguistic situation.

26. For a description of this tape-splicing technique, see Sec. C.1.

27. This conclusion seems to be in line with thoughts expressed by M. Haas in a conversation with me.

28. Including that of traditional Thai grammar. Thai orthography gives a consistent and easily read morphophonemic rendering of the five tones. See M. Haas, The Thai System of Writing (Washington, D. C., 1956), chap. III.

29. As does Henderson, Asia Major, N.S.I, 199, who, however, posits a "neutral" tone which is a one-unit tone system that contrasts with the usual five-unit system. See K. Pike, Tone Languages (Ann Arbor, 1948), p. 26, for "unstable neutral syllables."

30. Trager, p. 27.

31. Only five or six short studies published in the last couple of decades, plus two or three early and less reliable studies, are known to the writer. This deficiency will be largely remedied by J. Marvin Brown in a forthcoming Cornell University Ph. D. dissertation.

32. Falling terminal juncture is taken as the norm.

33. With /ʔ/ in her inventory, M. Haas can give a more elegant statement of distribution in terms of "smooth finals" and "stopped finals". See A Brief Description, p. 4.

34. See Tones (Sec. b), however, for neutralization of the static tonal features. Concomitant with this is the neutralization of the mid-central and low central vocalic features (/ə/: /a/), written as /a/ for morphophonemic convenience.

35. Thai informants show high agreement on the segmenting of utterances into words, including compound words. Doubtful cases were settled by testing potential word boundaries for the possibility of unlimited insertions, as proposed by Joseph H. Greenberg in Essays in Linguistics (Chicago, 1957), chap. 2.

36. Although it is true that the speaker may pause at any syllable boundary, especially in the isolative style.

37. See K. Pike, Tone Languages, pp. 16-17, 82-86.

38. It seems that in languages in which tones are retricted to certain syllables, analysis of intonation is easier. See Eunice V. Pike, "Tonemic-Intonemic Correlation in Mazahua (Otomi)," IJAL, XVII (1951), 37-41.

39. Søren Egerod, The Lungtu Dialect: A Descriptive and Historical Study of a South Chinese Idiom (Copenhagen, 1956), p. 51.

40. "Tones and Intonation in the Chengtu Dialect (Szechuan, China)," Phonetica, II (1958), 82-83.

41. Kruatrachue, a native speaker of Thai, comments, p. 57, ". . . it is difficult to determine whether there are distinctive elements in Thai stress, juncture, and intonation patterns." She then summarizes the statements of authors who do find distinctive elements.

42. These and other such utterances were elicited from two Thai informants in New York. At least one other Thai, a student of linguistics, casts doubt on the authenticity of this juncture, suggesting that it comes from English interference. This is of no relevance for the present study.

43. Spoken Thai, p. 282.

44. Asia Major, N.S.I, 204-208.

45. J. v. Laziczius, "Probleme der Phonologie," Ungarische Jahrbücher, XV (1935), cited by N. S. Trubetzkoy, Grundzüge der Phonologie (Prague, 1939), pp. 26-27.

46. "Techniques of Intensifying in Thai," Word, II (1946), 127-130.

47. Ellamac, Inc., Chicago, Ill.

48. R. K. Potter, G. A. Kopp, H. C. Green, Visible Speech (New York, 1947), chap. 2 and bibl.

49. Kay Electric Co., Pine Brook, N. J.

50. Franklin S. Cooper, "Spectrum Analysis," JASA, XXII (1950), 761-762; F. S. Cooper, "Some Instrumental Aids to Research on Speech," Fourth Annual Round Table Meeting on Linguistics and Language Teaching (Washington, D. C., 1954), pp. 48-50; John M. Borst, "The Use of Spectrograms for Speech Analysis and Synthesis," J. of the Audio-Engineering Soc., IV (1956), 14-15.

51. Original spectrograms made on a special photographic spectrograph can be played back through the transmission rather than reflection of light. This was not done here.

52. Homer Dudley, "Remaking Speech," JASA, XI (1939), 169-175.

53. John M. Borst and Franklin S. Cooper, "Speech Research Devices Based on a Channel Vocoder," JASA, XXIX (1957), 777 (A).

I. FORMANT FREQUENCIES OF THE VOWELS

The phonology of Standard Thai outlined in the Introduction posits nine vocalic phonemes which occur as single vowels and double vowels, the latter comprising geminates and vocalic clusters. Formant frequencies were obtained for these vowels by means of spectrographic measurements of isolated single and geminate vowels, vocalic clusters, and vowels in running discourse. Since the productions of only two speakers were measured, one set of vowels was checked for generality by exposure to Thai subjects in a listening test.

Two-formant vowels were synthesized and then tested on Thai subjects until optimum formant-frequency values were achieved. One reason for these experiments was that the bulk of evidence and thinking indicates that the first two formants contain most of the information needed for the identification of vowels. Another reason was the desire to arrive at a set of numerical values containing this information. Good support was given to the premise that the first two formants contain most of the relevant information; thus, the two-formant synthetic Thai vowels were highly intelligible and showed good agreement with real speech in their formant-frequency values.

A. Measurements

1. Formant Frequencies of Citation Forms of Single and Geminate Vowels

Spectrograms were made of all nine vowels, single and geminate, pronounced by W. N. and E. N. The vowels were recorded as citation forms in order to have steady-state formants without consonant-vowel transitions. Geminate vowels were uttered on all five tones and single vowels, on the high and low tones (Introduction B-3d).[1] Such utterances are quite normal in Thai speech, since the traditional name of a written vowel sign is a short utterance identical with the vowel. Indeed, children learning to read are made to recite such sequences as /aa àa âa áa ǎa/.[2] These recordings, as well as all others made for this study, were checked for normalcy by the original speakers and by at least one other Thai.

Since formants above the third seem not to contribute much to the identification of vowels,[3] the centers of the first three formants of each vowel were measured (Introd. C.1). All steady-state vowels were measured at a point halfway through their duration. Wherever formants moved up or down, measurements were taken at enough points to show the nature and extent of the deviation from the steady-state. Such points are marked in the tables as percentages of the total duration of the vowel.

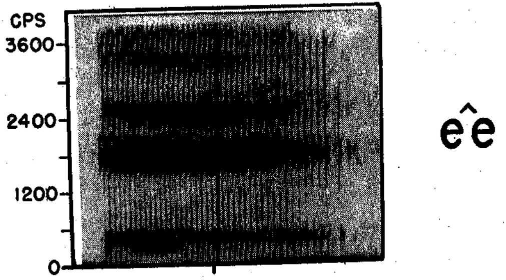
Precision of measurements

All measurements of formant frequencies were made as carefully as possible within the limits set by uncertainty as to the centers of formants, which had to be found by eye. The formant locations on the frequency scale, measured in inches and converted to frequencies, usually gave numbers such, for example, as 461, 846 and 2652 cps for the formants of W. N.'s /ò/, which imply greater accuracy than is in fact warranted.

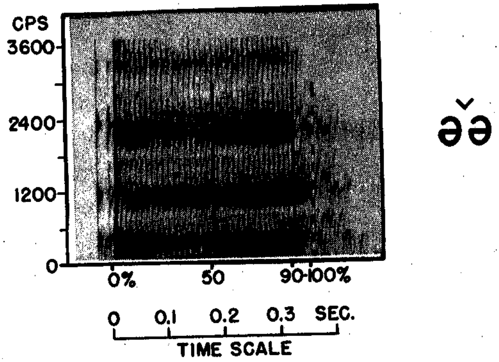
An experiment was run to establish the standard deviation of the measurements. A sheet of tracing paper was laid over a spectrogram of W. N.'s production of /ee/, and lines were drawn by eye along the centers of the first three formants through a line constructed perpendicular to the base halfway through the vowel, as shown in Figure 1.1A. The first formant was clearly defined, the others, a little less so. The three points at the intersections of the lines were measured against the frequency scale of the spectrogram. This was done ten times on ten separate pieces of paper. The standard deviation of the ten measurements of each formant was then calculated by the formula $s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$ in which the x_i's are the measurements, $\bar{x}$ is the mean of the measurements, and $n=10$, the number of measurements.[4]

The standard deviations were 10 cps for the first formant, 18 cps for the second, and 19 cps for the third. Since these three formants seemed fairly representative of the varying degrees of sharpness of definition of formants in the spectrograms made for this study, it was decided to take the lowest of the three standard deviations as a guide in rounding the measured values to numbers which give a realistic indication of the attainable precision. Thus, all measurements of formant frequencies have been rounded off to the nearest 10 cps.

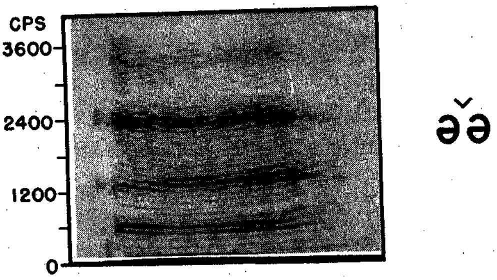
The question of precision is also involved in the choice of points in time at which to measure moving formants. The points



A. FORMANT CENTERS USED FOR CALCULATION OF STANDARD DEVIATION. SUBJECT W.N.



B. POINTS AT WHICH FORMANT FREQUENCIES WERE MEASURED. SUBJECT E.N.



C. NARROW-BAND SPECTROGRAM. SUBJECT E.N.

FIG. 1.1

at which frequency measurements were made are given as percentages of the vowel durations. This sometimes requires a slightly arbitrary decision as to the absolute limits of the vowel. What is wanted is a reasonable criterion for a measure of the duration of what is heard as a vowel.

In the absence of a physical criterion amenable to rigorous definition and one that provides a sure measure of perceived duration, a reasonable criterion is prominence of at least one of the first three formants. This prominence correlates well with periodic glottal excitation, which is manifested in wide-band spectrograms (Fig. 1.1A, B) by regularly spaced vertical striations and in narrow-band spectrograms (Fig. 1.1C) by the presence of harmonic traces, more or less in the horizontal plane. In the ideal case, vibration of the vocal cords starts at the beginning of the vowel and comes neatly to a halt at the end of the vowel. The practical case is not always so simple. The periodic portion of the vowel may merge into a stretch of quasi-periodic excitation such as appears before the point marked 100% in Figure 1.1B. As long as this quasi-periodic portion is intense enough to be an obviously integral continuation of formants, it is taken as contributing to the duration of the vowel.

Although it was usually easy to mark the ends of a vowel by means of the criterion of formant prominence, there were two more kinds of indeterminacy in some of the Thai spectrograms: (1) when a glottal stop occurs just before or after a vowel, one or more glottal pulses, sometimes accompanied by aperiodic noise, can be seen separated from the main bodies of the formants. This appears after the formants in Figure 1.1A and before the formants in Figure 1.1B. (2) In some of the vowel citations, the speaker must have stopped phonating but held his articulators in position for a moment before closing his mouth or starting the next utterance. When this happened, a breathy release was sometimes evident as a weak aperiodic continuation of the formants. Some indication of this can be seen beyond the 100% mark in Figure 1.1B. Wherever they were distinguishable, these two features were excluded from the measurements of vowel duration.

In any vowel with moving formants, then, the duration of the vowel was measured as accurately as possible; nevertheless, a certain amount of arbitrariness could not be avoided. In the worst cases, a boundary mark might have been moved 20 milliseconds to the right or left; however, such a discrepancy was not likely to introduce a difference of more than 80 cps in the initial and final values of the formants.

Since it was a matter of interest in this study to tabulate formant frequencies in sets of three, points were chosen for obtaining these values only where all three formants were prominent. Wherever the last point entered in a table is greater than 50% but less than 100% of the duration, this means that a set of three values could not be obtained beyond that point either because one or two formants were no longer evident, or because at least one formant was manifested by weak quasi-periodic or aperiodic sound that made finding the formant center impossible. On the other hand, if the points given extend from the 50% mark through one half of the vowel, this means that the other half is in a steady-state at the formant frequencies of the 50% point.

The spectrogram in Figure 1.1B illustrates the conventions that have been used. The vertical lines marked 0% and 100% delimit the duration of the vowel. The vowel is nearly steady-state, but the slight movement of the formants has been indicated by measurements at three points, the beginning and middle of the vowel and at 90% of the duration, the last point at which all three formant centers could be measured with confidence.

The formant frequencies are given in separate tables for the two informants. There are double entries for W. N.'s productions of single vowels on the low tone (F. N. 1), but it should be understood that no two repetitions of any vowel will necessarily have exactly the same formant frequencies.

TABLE 1.1

Informant: W. N.

Vowel	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ì		350	2050	2530
(h) ì		380	2080	2730
(h) í		380	2150	2810
ii		260	2220	2790
ìi		300	2190	2750
íi		260	2260	2790
îi	33%	300	2260	2860
	100	380	2000	2600
ǐi		300	2260	2830

TABLE 1.1 — Continued

Vowel	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
è		440	1980	2420
(h) è		500	2000	2690
(h) é		580	1920	2570
ee		450	2000	2640
eè		410	2070	2560
eé		530	1960	2640
eê		520	1970	2620
eě		450	1990	2620
æ̀	0%	700	1800	2490
	50	700	1720	2420
	100	700	1690	2310
(h) æ̀	0	690	1770	2340
	50	690	1690	2340
(h) ǽ		730	1700	2310
ææ		670	1840	2440
ææ̀		670	1840	2470
æǽ	50	670	1800	2470
	100	750	1700	2440
ææ̂		680	1760	2290
ææ̌	0	640	1910	2510
	50	670	1800	2470
	100	750	1690	2400
ɨ̀		460	1190	2460
(h) ɨ̀		460	1270	2460
(h) ɨ́		420	1270	2420
ɨɨ		490	1420	2510
ɨɨ̀		450	1390	2360
ɨɨ́		450	1390	2590
ɨɨ̂	50	450	1350	2590
	100	450	1390	2440
ɨɨ̌		410	1350	2550
ə̀	50	540	1130	2560
	100	540	1210	2610
(h) ə̀		540	1190	2540
(h) ə́		560	1190	2540

TABLE 1.1 — Continued

Vowel	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
əə		560	1200	2470
ə̀ə		520	1240	2400
əə́		560	1260	2590
ə̂ə	50%	600	1310	2470
	100	540	1270	2400
ə̌ə		490	1240	2400
à		770	1230	2460
(h) à	50	690	1230	2540
	100	770	1230	2460
(h) á		710	1250	2690
aa		710	1160	2360
àa		710	1140	2320
aá	50	710	1160	2250
	100	710	1160	2380
âa		730	1200	2320
aǎ	50	710	1160	2250
	100	710	1200	2320
ù		380	690	2610
(h) ù		350	670	2460
(h) ú		350	710	2610
uu		340	670	2510
ùu		370	670	2510
uú		370	750	2550
uû	0	390	860	2550
	50	410	790	2490
uǔ		390	710	2510
ò		460	850	2650
(h) ò	0	460	810	2650
	50	460	810	2540
(h) ó		500	880	2650
oo		450	860	2490
òo		490	840	2510
oó		490	860	2550
oô	50	490	900	2470
	100	490	860	2400
ǒo		480	810	2420

TABLE 1.1 — Continued

Vowel	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ɔ̀	50%	600	980	2610
	100	600	980	2540
(h) ɔ̀	0	610	1000	2560
	50	610	1000	2460
	100	610	1000	2420
(h) ɔ́		650	1000	2570
ɔɔ		590	950	2350
ɔɔ̀		570	900	2350
ɔɔ́	0	590	990	2350
	50	590	990	2200
ɔɔ̂		590	990	2240
ɔɔ̌	0	590	920	2400
	50	590	920	2270
	100	590	1030	2270

TABLE 1.2

Informant: E. N.

Vowel	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ì		350	2120	2980
í		380	2160	2930
ii		330	2260	3130
iì		320	2240	3130
ií		310	2210	3090
iî	0%	280	2260	2400
	50	280	2320	3190
	80	290	2260	3060
iǐ		330	2260	3120
è		410	1860	2610
é		420	1860	2570
ee		410	2050	2530
eè		400	2040	2400
eé		410	2010	2400

TABLE 1.2 — Continued

Vowel	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
eê		430	2020	2400
eě	50%	410	2040	2400
	100	450	2000	2380
æ̀		610	1620	2360
ǽ		570	1660	2320
ææ		640	1840	2460
æ̀æ	50	600	1880	2460
	75	640	2010	2530
	100	600	1930	2460
ǽæ		600	1860	2460
æ̂æ		640	1930	2530
æ̌æ		640	1880	2460
ɨ̀		340	1330	2590
ɨ́		380	1380	2650
ɨɨ		400	1470	2640
ɨ̀ɨ		350	1470	2490
ɨ́ɨ		380	1470	2690
ɨ̂ɨ	0	400	1490	2530
	50	400	1520	2660
	100	400	1490	2600
ɨ̌ɨ		390	1470	2500
ə̀		510	1240	2610
ə́		540	1240	2520
əə		480	1310	2600
ə̀ə	0	510	1280	2520
	50	380	1330	2330
ə́ə		530	1350	2660
ə̂ə	0	550	1400	2470
	50	510	1330	2620
ə̌ə	0	530	1240	2420
	50	470	1330	2460
	90	500	1400	2500
à		660	1270	2400
á	0	660	1360	2370
	50	660	1400	2370

TABLE 1.2 — Continued

Vowel	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
aa		650	1400	2330
àa		650	1400	2330
áa	0%	670	1450	2330
	50	730	1530	2400
âa		670	1470	2370
ǎa		670	1360	2370
	90	730	1530	2330
ù		370	680	2520
ú		340	830	2640
uu		370	830	2680
ùu		400	830	2530
úu	0	310	670	2530
	50	310	670	2640
	90	310	670	2800
uû	50	360	810	2660
	90	350	730	2600
uǔ	50	340	670	2400
	100	400	800	2530
ò		500	830	2560
ó		530	940	2660
oo		450	960	2650
òo		470	960	2540
oó	0	470	870	2840
	50	540	920	2810
	100	540	930	2730
ôo	0	480	960	2610
	50	470	920	2600
ǒo	50	470	870	2630
	90	530	930	2730
ɔ̀		610	940	2720
ɔ́		630	1040	2560
ɔɔ		640	1070	2540
ɔ̀ɔ	0	630	1070	2530
	50	630	930	2440
ɔɔ́	0	630	1070	2650
	50	630	1070	2460

TABLE 1.2 — Continued

Vowel	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ɔ̂ɔ	0%	600	1070	2610
	50	600	1010	2480
	100	600	930	2600
ɔ̌ɔ	0	600	1000	2400
	50	600	910	2260
	90	670	1070	2400

Wide-band spectrograms of all the vowels in Table 1.2 are given in Figures 1.2-8. Each figure shows all nine of E. N.'s single or geminate vowels on one tone.

2. Vowel Plots

The formant data were viewed in another way by plotting the first formant against the second logarithmically.[5] Such a display, with the frequency of the first formant increasing downward along the Y-axis and the second increasing to the left along the X-axis, looks very much like a traditional vowel quadrilateral. A logarithmic plot was made for the vowels of each informant on each tone.

To arrive at values for two-formant synthetic vowels for the experiments on perception, the plots were superimposed and an average of the scatter values of each vowel was taken. This was done separately for geminate vowels and single vowels. For example, the single-vowel values of Figure 1.10 were derived from the scatter plots of Figure 1.9.[6] The scatter of these points is too large to be accounted for by imprecision in measurement or calibration; nevertheless, in none of the charts did the scatter plots of any two vowels overlap. Figure 1.10 shows the average formant frequencies of W. N.'s vowels, and Figure 1.11, E. N.'s vowels. Figure 1.12 provides a direct comparison of the single vowels of the two informants, and Figure 1.13, the geminate vowels.

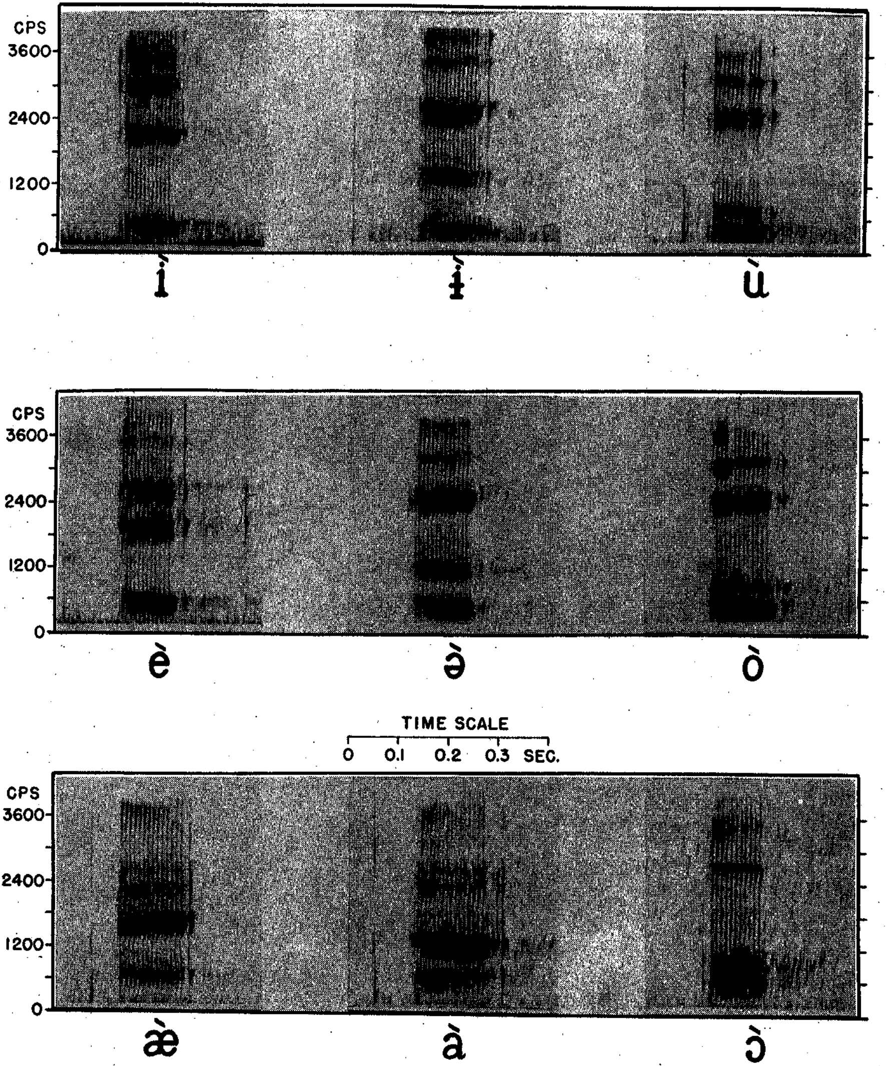


FIG. 1.2 SPECTROGRAMS OF E.N.'S SINGLE VOWELS ON THE LOW TONE

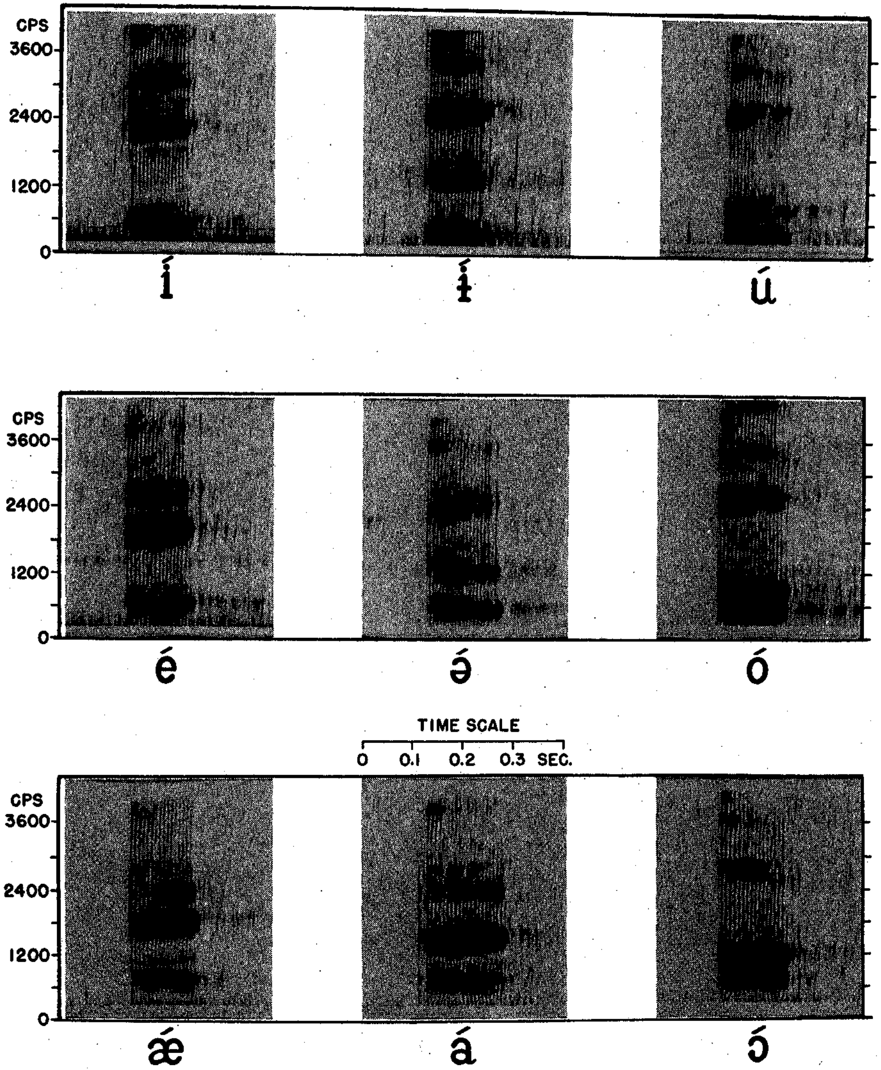


FIG. 1.3 SPECTROGRAMS OF E.N.'S SINGLE VOWELS ON THE HIGH TONE

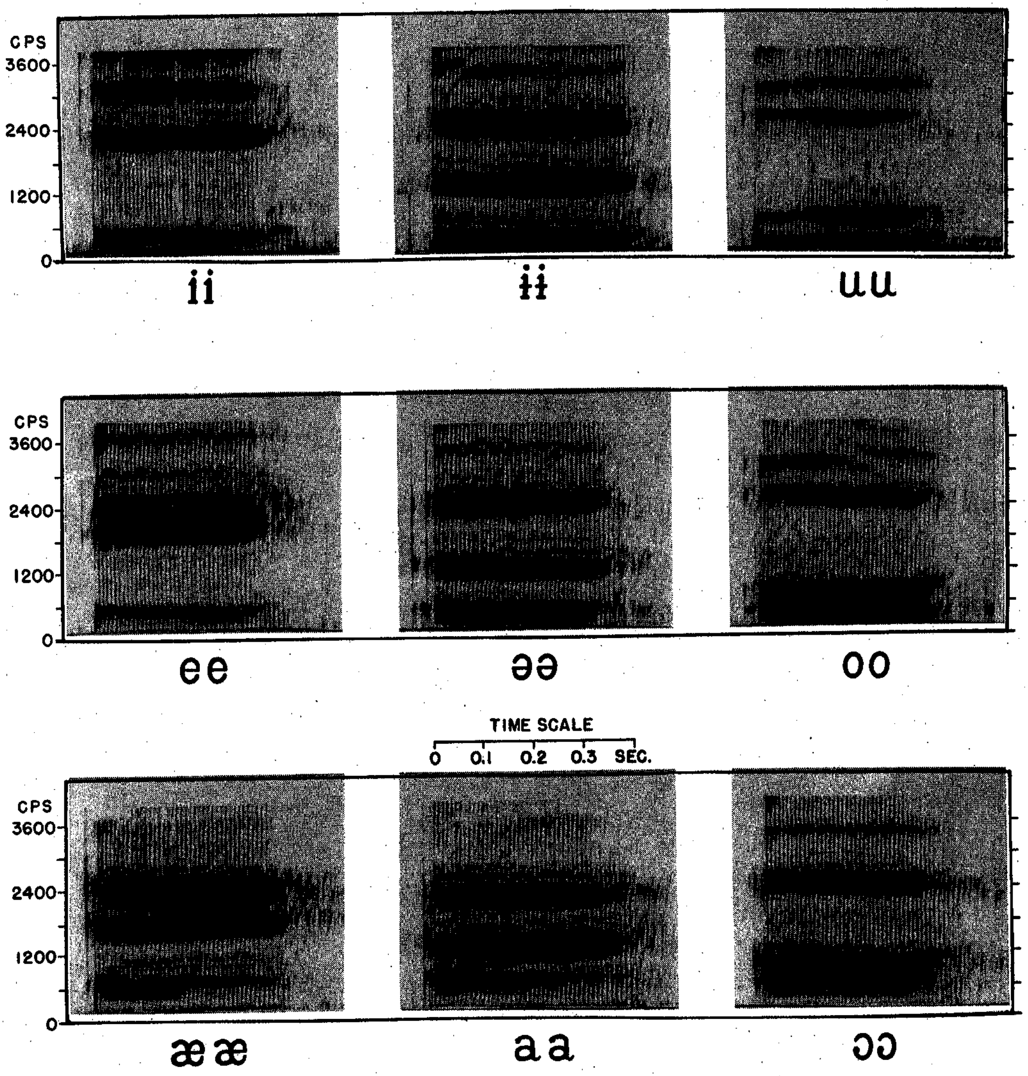


FIG. 1.4 SPECTROGRAMS OF E.N.'S GEMINATE VOWELS ON THE MID TONE

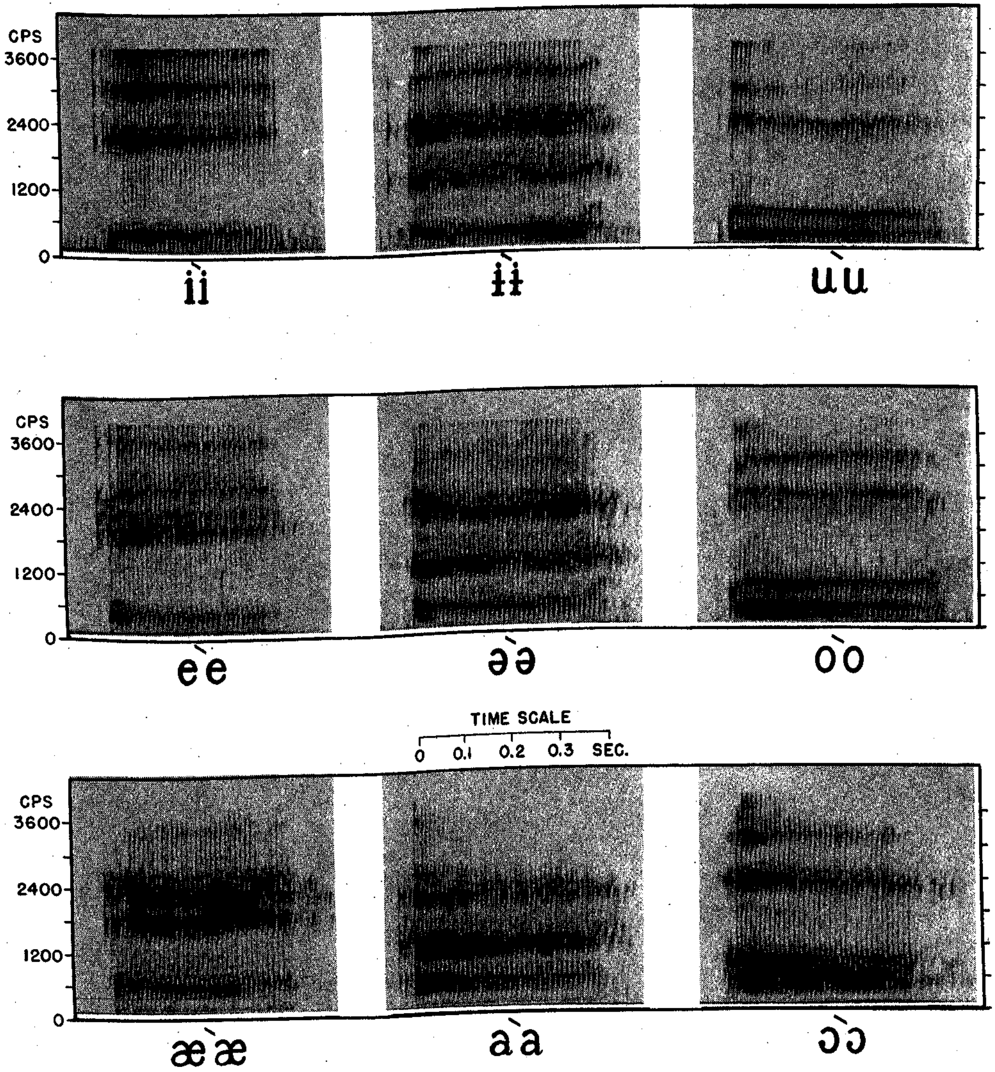


FIG. 1.5 SPECTROGRAMS OF E.N.'S GEMINATE VOWELS ON THE LOW TONE

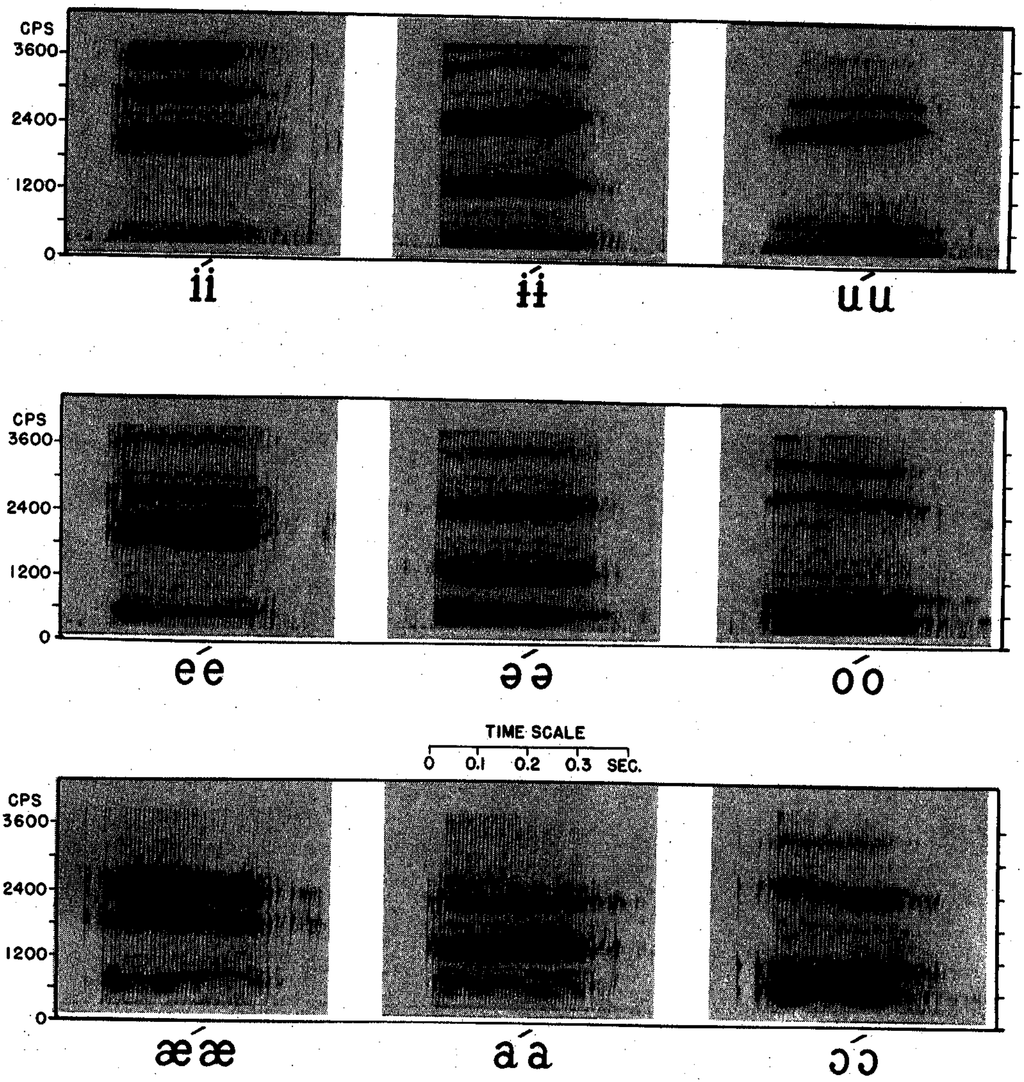


FIG. 1.6 SPECTROGRAMS OF E.N.'S GEMINATE VOWELS ON THE HIGH TONE

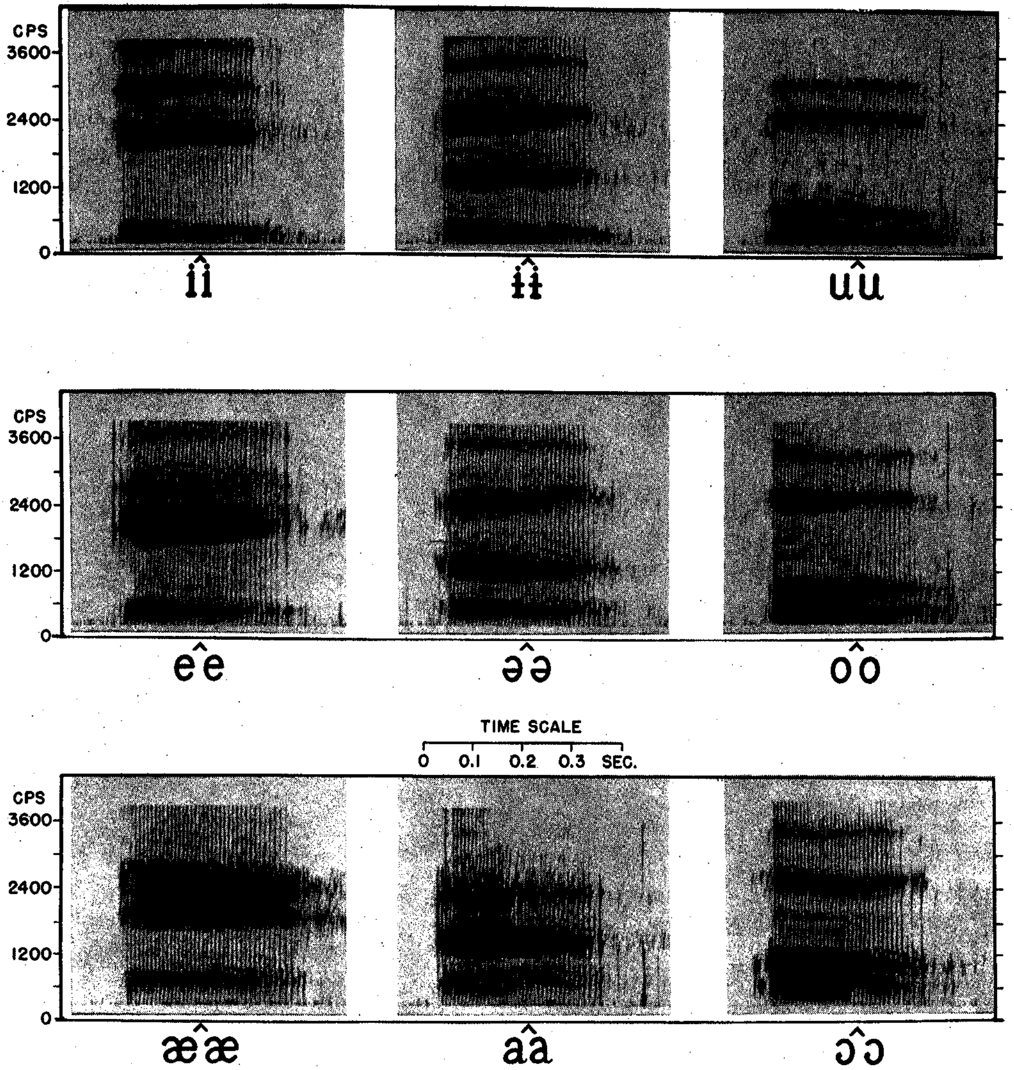


FIG. 1.7 SPECTROGRAMS OF E.N.'S GEMINATE VOWELS ON THE FALLING TONE

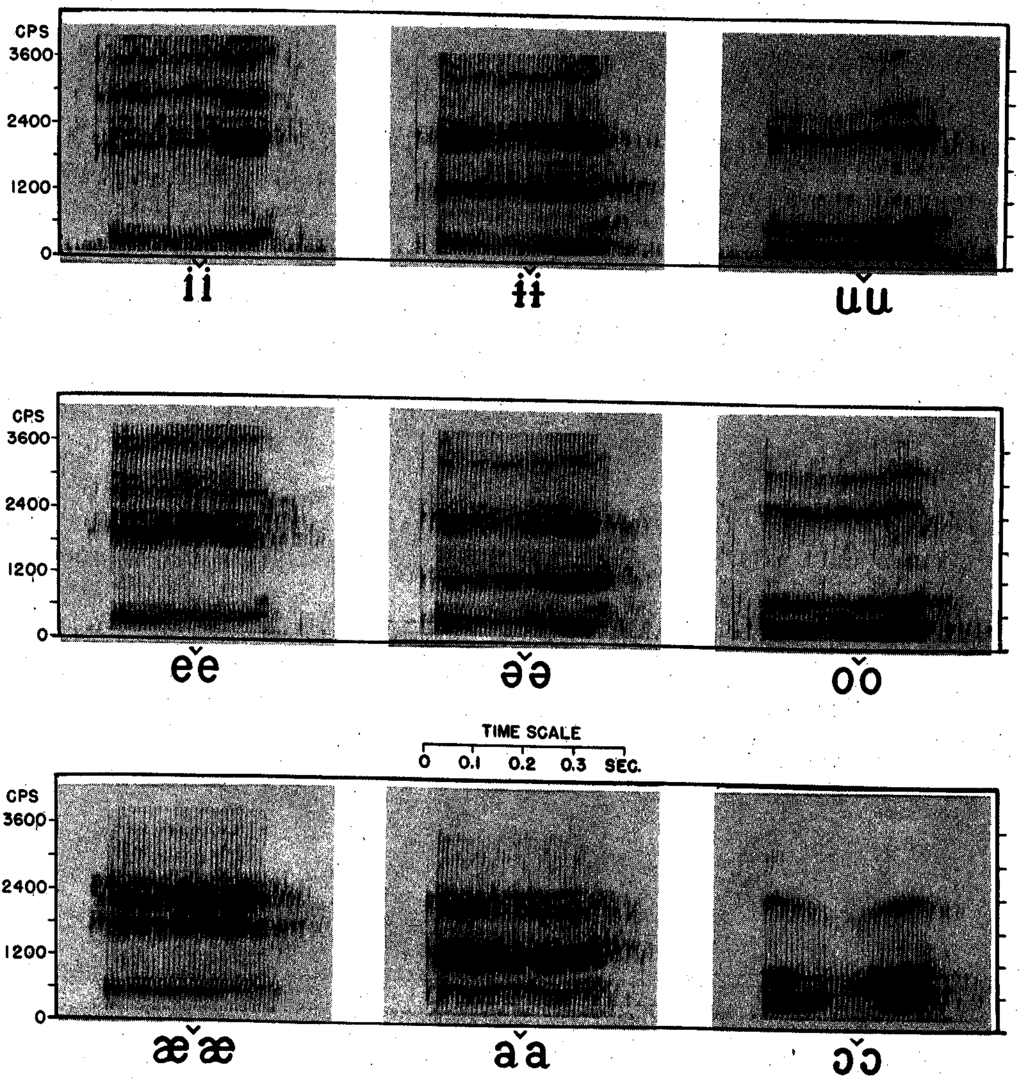


FIG. 1.8 SPECTROGRAMS OF E.N.'S GEMINATE VOWELS ON THE RISING TONE

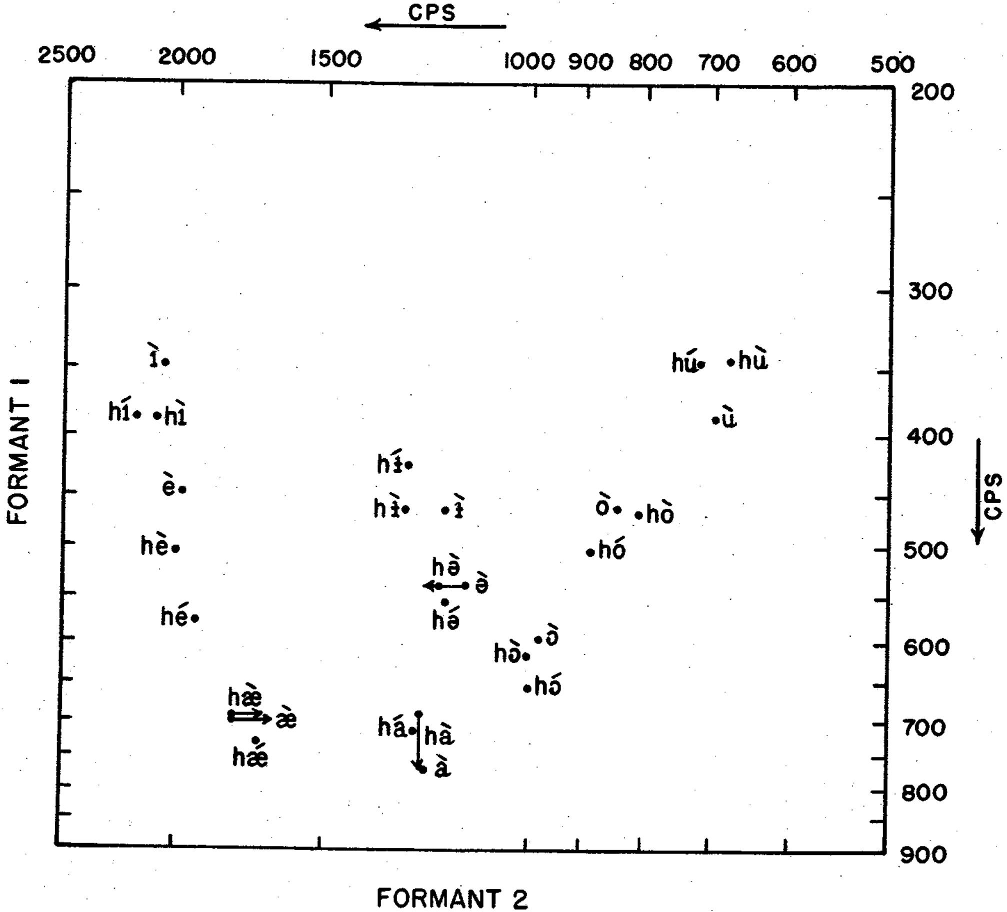


FIG. 1.9 SCATTER PLOT OF W.N.'S SINGLE VOWELS

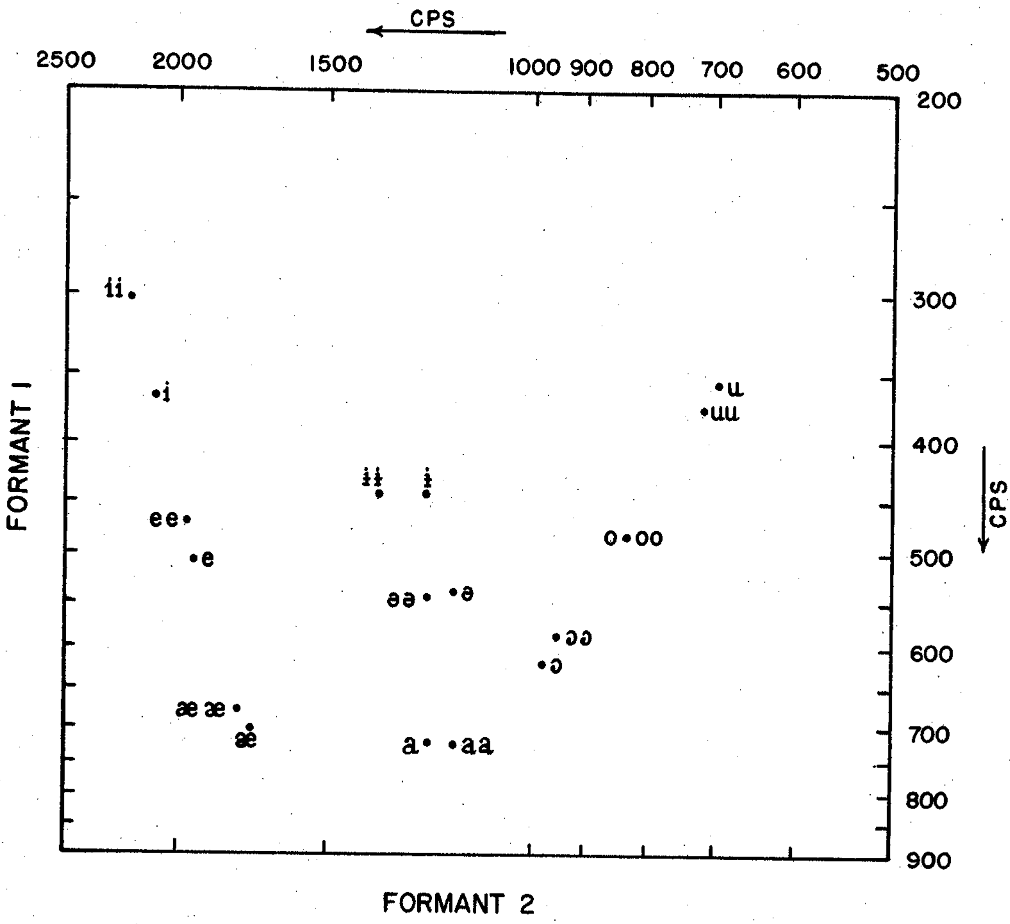


FIG. 1.10 AVERAGE FORMANT FREQUENCIES OF W.N.'S VOWELS

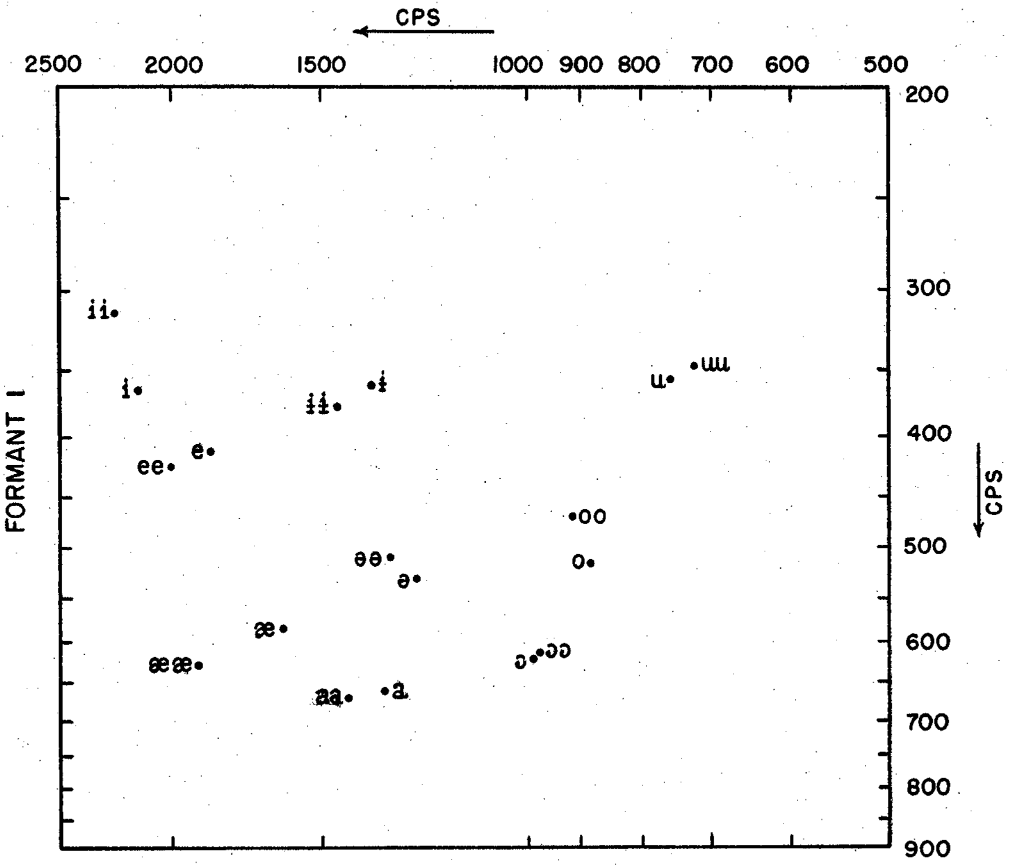


FIG.I.II AVERAGE FORMANT FREQUENCIES OF E.N.'S VOWELS

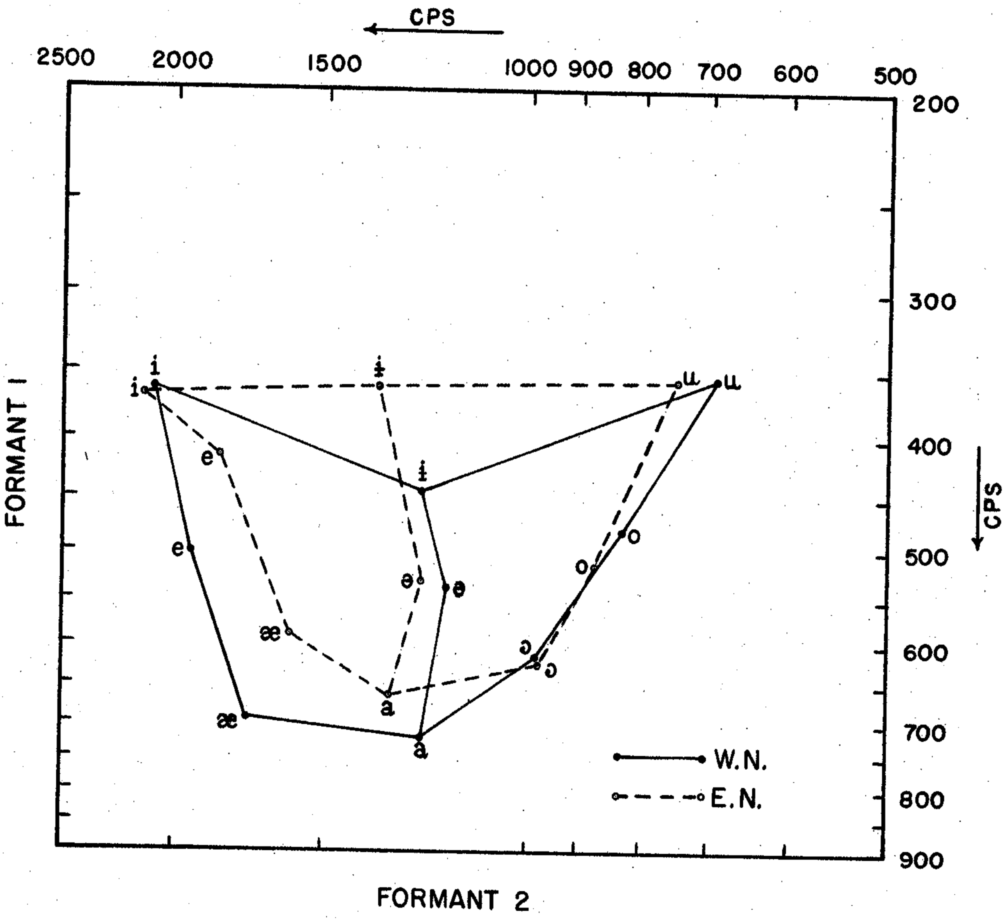


FIG. 1.12 COMPARISON OF SINGLE VOWELS IN FIGURES 10 AND 11

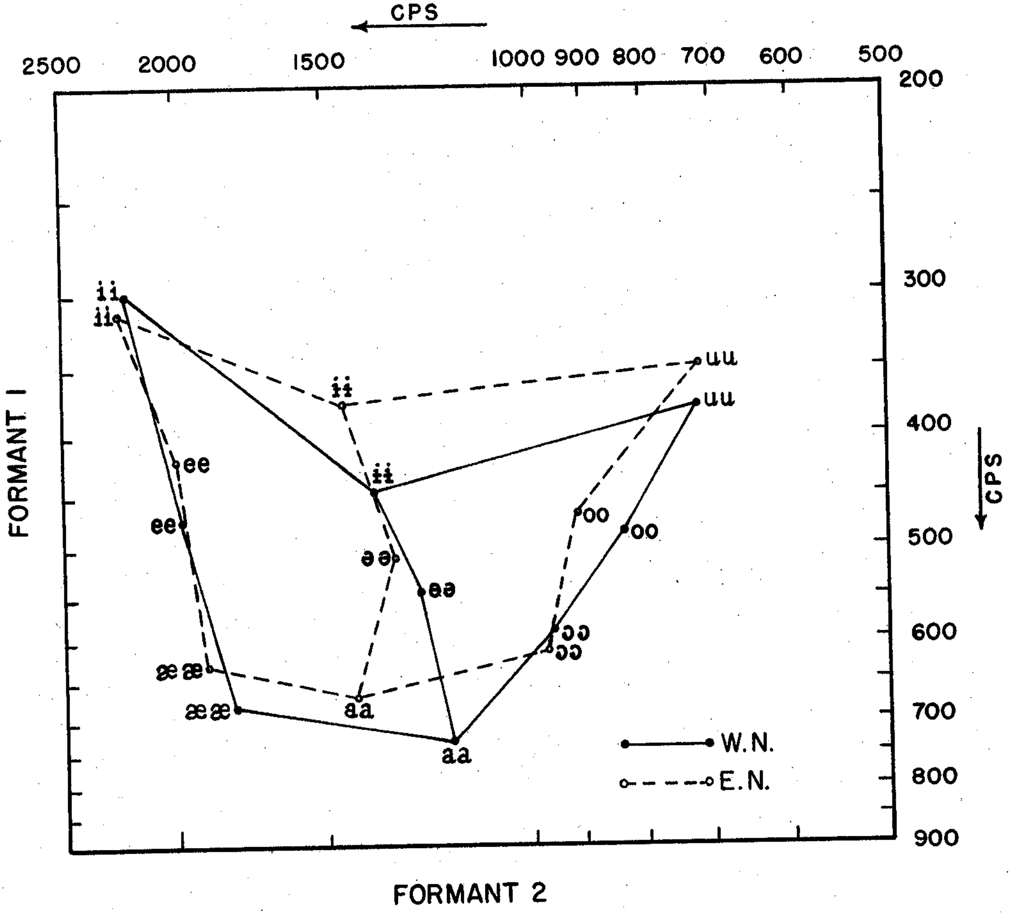


FIG. 1.13 COMPARISON OF GEMINATE VOWELS IN FIGURES 10 AND 11

3. A Sampling of Vowels in Connected Speech

a. Vocalic clusters

As a transitional step from vowels in isolation to vowels in running discourse, spectrograms were made of the vocalic clusters /ia ɨa ua/ spoken by E. N. on all five tones as citation forms. Spectrograms of the clusters on the mid tone are shown in Figure 1.14. The formant frequencies are entered in Table 1.3 according to the same conventions as in Tables 1.1-2.

TABLE 1.3

Informant: E. N.

Cluster	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ia	4%	300	2150	3050
	40	310	2140	3010
	44	370	2510	2920
	48	410	1990	2780
	52	510	1850	2650
	56	540	1710	2560
	60	610	1530	2450
	64	630	1460	2370
	84	670	1460	2300
ìa	7	290	3160	3110
	40	280	2110	2980
	45	370	2060	2890
	50	410	1990	2780
	55	510	1860	2650
	60	550	1700	2550
	65	590	1600	2510
	70	630	1480	2390
	75	640	1460	2320
	85	640	1460	2320
ía	55	310	2110	2990
	60	370	1990	2880
	65	410	1970	2760
	70	530	1860	2650
	75	540	1730	2520
	80	570	1660	2410

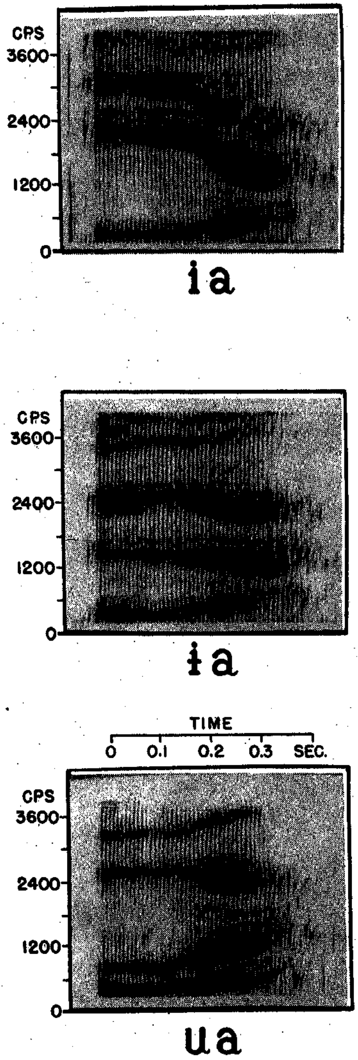


FIG. 1.14 SPECTROGRAMS OF THE THREE VOCALIC CLUSTERS ON THE MID TONE

TABLE 1.3 — Continued

Cluster	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
	90%	610	1570	2260
	100	610	1520	2260
iâ	0	300	2100	3020
	50	340	2120	2980
	56	370	2100	2920
	67	470	1890	2680
	78	550	1660	2590
iǎ	9	310	1990	2980
	32	310	2060	3050
	45	310	1990	2920
	50	370	1980	2780
	55	410	1860	2650
	60	490	1820	2520
	65	540	1730	2480
	70	570	1620	2400
	75	670	1600	2390
	80	700	1530	2320
ɨa	4	370	1460	2520
	36	340	1460	2550
	40	380	1460	2550
	44	410	1460	2520
	48	430	1450	2520
	52	530	1460	2450
	56	540	1440	2390
	60	550	1370	2370
	64	590	1330	2370
	78	670	1330	2360
	88	670	1460	2230
ɨà	0	380	1400	2490
	32	350	1450	2370
	40	410	1450	2360
	44	470	1450	2350
	52	550	1340	2320
	72	570	1330	2310
	84	660	1460	2270

TABLE 1.3 — Continued

Cluster	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ɨ́a	0%	370	1450	2520
	42	380	1440	2640
	53	410	1460	2650
	58	430	1460	2610
	64	470	1460	2610
	73	540	1460	2390
	78	550	1460	2390
	100	550	1460	2390
ɨ̂a	10	340	1410	2550
	47	340	1420	2650
	51	410	1420	2620
	56	500	1450	2490
	60	530	1450	2390
	65	530	1450	2320
	82	540	1460	2320
ɨ̌a	40	340	1400	2390
	48	410	1420	2390
	52	470	1440	2360
	56	540	1440	2260
	60	570	1440	2260
	84	570	1440	2260
ua	22	340	720	2520
	30	350	750	2520
	39	380	830	2520
	43	410	940	2490
	47	490	1040	2470
	52	530	1100	2440
	56	540	1200	2420
	65	550	1210	2400
	73	550	1330	2390
	90	550	1340	2390
ùa	27	410	670	2520
	31.5	410	680	2510
	36	410	740	2400
	40.5	420	800	2370
	45	470	870	2360
	49.5	510	1000	2320

TABLE 1.3 — Continued

Cluster	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
	54%	540	1070	2280
	58.5	470	1200	2280
	81	470	1330	2320
uá	24	340	800	2640
	36	380	820	2600
	42	390	910	2560
	48	420	950	2550
	54	530	1070	2520
	60	540	1200	2510
	72	550	1270	2390
	96	550	1310	2390
uâ	0	410	800	2520
	27	390	790	2560
	31.5	410	830	2550
	36	510	940	2550
	40.5	530	1040	2520
	45	540	1070	2490
	49.5	550	1200	2410
	54	570	1210	2390
	63	550	1250	2360
	72	550	1270	2340
	81	540	1310	2320
ǔa	0	390	760	2410
	36	390	790	2390
	40.5	410	820	2390
	45	420	950	2370
	49.5	490	1070	2320
	54	540	1090	2320
	58.5	540	1200	2320
	67.5	570	1230	2320
	76.5	590	1310	2320

b. Vowels in running discourse

For this part of the analysis, about half an hour of unrehearsed narration was recorded by E. N. This consisted of personal anecdotes, folktales and glimpses of Thai history. The speech was

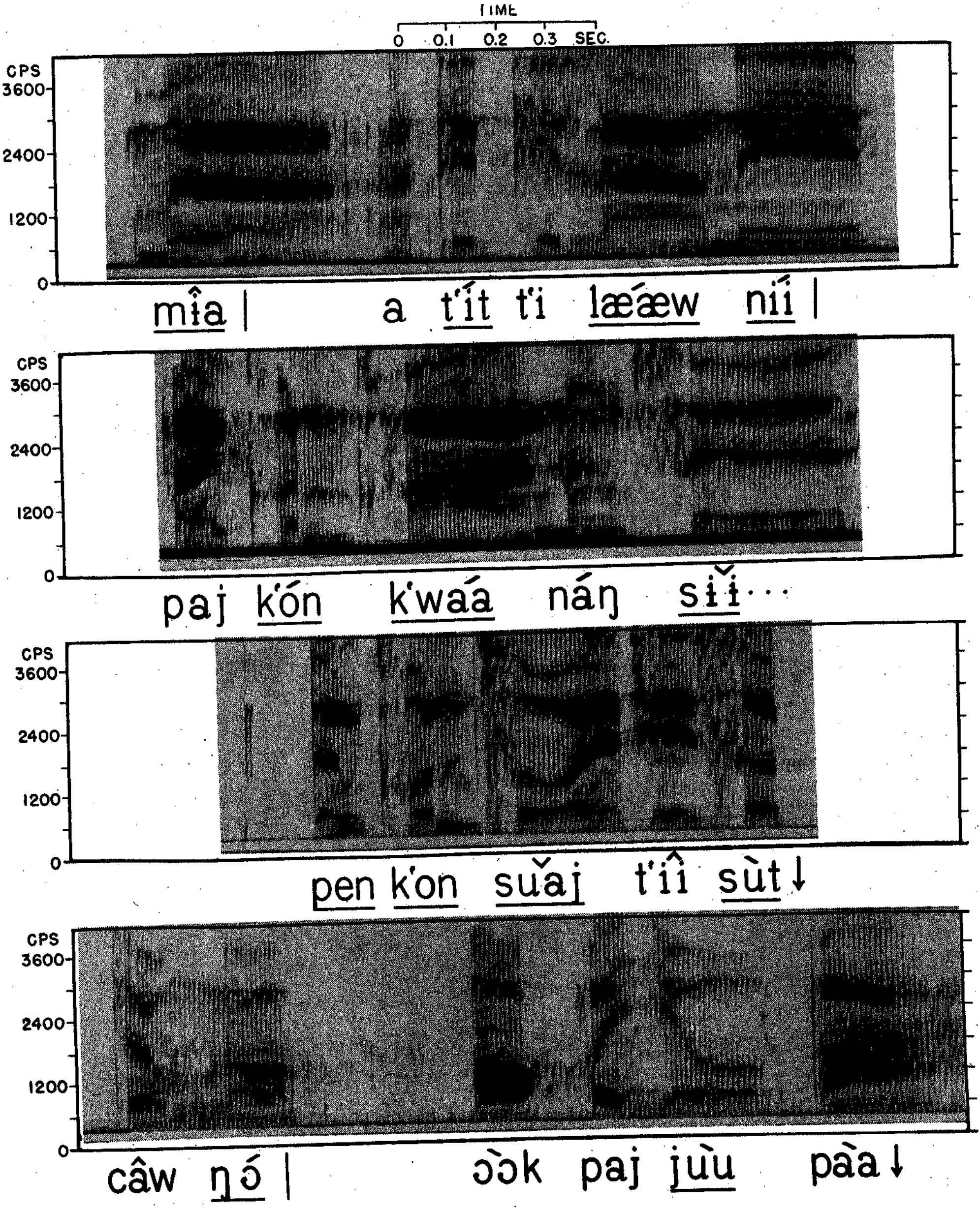


FIG. 1.15 VOWELS IN RUNNING SPEECH

judged to be of a normal colloquial type in the combinative style with occasional stretches in the rapid combinative style (Introduction B-1). Here and there a word is spoken in the isolative style. The recording was long enough to yield each vowel in at least one phonemic environment. Examining each vowel in its total distribution[7] would have been far too ambitious an undertaking as well as unnecessary for the purpose at hand. It was deemed sufficient, instead, to trace the formants of what seemed to be a representative sample of vowels in connected speech. That is, twenty-four spectrograms of sentences and fragments of sentences were made to show instances of each vowel in as many types of phonemic environment as could be found in the recording. Wherever only one environment was found for a single or geminate vowel, there was at least one more for its geminate or single counterpart. Four of these spectrograms are shown in Figure 1.15. Syllables underlined in the transcription are entered in the formant table (Table 1.4).

Table 1.4 gives an overall view of the allophonic variation of the vowels in physical terms. This information could serve as a point of departure in a future study of the consonants, in which data on consonant-vowel transitions would be crucial. Points to be measured were chosen in the same way as for the citation forms, although it was seldom possible here to choose just one point for a given vowel. Moreover, it should be understood that points chosen at or near the ends of formants may be in consonant-vowel transitions and consequently just as much a part of the consonant as the vowel. The semivowels /w j/ are included in the table as extensions of the formants of neighboring vowels.

TABLE 1.4

E. N.'s Vowel Formants in Connected Speech

Vocalic Nucleus	Word	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ji	dâjjin ‘hear’	0%	170	1930	3010
		22	220	1970	3010
		66	290	1910	2650
		100	280	1790	2490
í	aat‘ít ‘week’		400	2060	2650

TABLE 1.4 — Continued

Vocalic Nucleus	Word	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ii	dii 'to be good'	50%	340	2190	2760
		100	340	2100	2590
ií	nií 'this'	0	400	1860	2520
		16	370	2110	2660
		24	370	2120	2740
		80	370	2120	2740
		100	370	1930	2590
ǐa	sǐacaj 'to be sorry'	0	370	1850	2580
		20	390	1960	2650
		40	400	2000	2690
		60	400	1980	2650
		80	370	1880	2540
		100	300	1840	2530
e	pen 'to be (something)'		470	1600	2560
é	lék 'to be small'		490	1860	2590
ee	ceedii 'stupa'		390	1930	2720
æ̀	pɨ̀kpʻæ̀n 'firmness'		790	1740	2410
æ̀	tæ̀ (~tæ̀æ) 'but'		650	1600	2570
æǽw	læǽw 'finished'	0	700	1600	2590
		9.5	790	1620	2590
		28.5	820	1630	2600
		38	820	1600	2560
		57	820	1460	2490
		66.5	800	1400	2450
		76	670	1330	2450
		100	540	1330	2450
æ̂æ	mæ̂æ 'mother'	0	670	1440	2520
		50	700	1660	2520
		100	490	1460	2520
ɨ̀	pɨ̀kpʻæ̀n 'firmness'		400	1400	2620
ɨ̌	tʻɨ̌ŋ 'to reach'		280	1740	2570
ɨ̌ɨ	nǎŋsɨ̌ɨ 'book'	50	300	1540	2410
		100	300	1540	2490

TABLE 1.4 — Continued

Vocalic Nucleus	Word	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
ɨ᷍a	mɨ᷍a ‘when’	0%	470	1340	2520
		5	470	1520	2530
		10	470	1600	2560
		15	470	1600	2570
		20	470	1600	2590
		35	470	1540	2530
		40	500	1500	2520
		60	710	1480	2510
		65	780	1460	2490
		100	780	1460	2490
ə	ŋən ‘money’		590	1610	2540
əə	dəən ‘to walk’		470	1540	2520
əə̀	bəə̀k ‘to withdraw (money)’	0	470	1200	2530
		50	520	1450	2620
		100	430	1470	2530
əəj	k‘əəj ‘ever’	0	260	1450	2410
		15	340	1530	2410
		30	430	1580	2390
		45	430	1690	2370
		60	430	1730	2360
		75	430	1730	2370
		90	410	1660	2390
		100	300	1600	2430
âw	kâw ‘nine’	0	470	1600	2350
		12.5	670	1570	2350
		25	800	1460	2350
		37.5	800	1330	2350
		50	740	1280	2350
		62.5	610	1230	2390
		75	570	1130	2450
		87.5	410	1040	2520
		100	410	1040	2520
a	samǎj ‘occasion’		540	1200	2680
a	ma (∼maa) ‘ago’		610	1360	2520

TABLE 1.4 — Continued

Vocalic Nucleus	Word	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
á	rák ‘to love’	0%	410	1460	2520
		50	580	1380	2490
		100	780	1330	2450
à	jàk (~jàak) ‘would like’	0	260	1930	2820
		17	340	1910	2810
		33	540	1790	2650
		50	740	1600	2520
		67	830	1560	2510
		84	870	1530	2510
		100	860	1520	2520
aa	naan ‘a long time’		790	1360	2520
àa	ookàat ‘opportunity’	0	510	1600	2450
		35	720	1480	2450
		60	910	1460	2490
		100	740	1570	2610
wáa	k‘ónk‘wáa ‘to do research’	16	740	1170	2480
		40	780	1310	2520
		80	780	1400	2450
		100	780	1460	2450
âa	mâak ‘to be much’		780	1460	2390
aa	baroo+mát‘âat ‘name of a monastery’	0	940	1640	2520
		25	740	1400	2430
		50	650	1380	2390
		100	650	1460	2520
ǎaj	lǎaj ‘several’	0	740	1400	2450
		50	780	1660	2490
		100	280	1860	2560
u	k‘un ‘you’		370	1030	2570
ù	sùt ‘most’	0	380	1360	2540
		50	380	1310	2520
		100	380	1270	2490
jùu	jùu ‘to dwell’	0	140	2120	2390
		10	160	1990	2450
		20	170	1700	2480

TABLE 1.4 — Continued

Vocalic Nucleus	Word	Point of Measurement	F-1 (cps)	F-2 (cps)	F-3 (cps)
		30%	180	1130	2490
		40	200	940	2520
		50	210	870	2530
		60	210	800	2570
		100	210	800	2570
ùa	k‘ùap ‘year (of a child's age)’	0	380	870	2550
		16	360	870	2550
		32	360	910	2550
		48	380	1000	2530
		64	420	1070	2510
		80	460	1080	2440
		100	410	1000	2410
o	k‘on ‘person’		410	940	2690
o	okàat (∼ookàat) ‘opportunity’		390	940	2680
ó	k‘ónk‘wáa ‘to do research’		390	1130	2520
ǒ	p‘ǒm ‘I (polite male)’		500	920	2650
oǒ	suk‘oǒt‘aj ‘Sukhothai’	0	410	1010	2650
		45	410	940	2650
		100	420	1200	2660
ɔ́	ŋɔ́ ‘negrito’		610	1070	2520
ɔ̂ɔ	kɔ̂ɔ ‘then’		500	1040	2370

B. Experiments on the Perception of Vowel Qualities

The goal at this stage of the work was to obtain a set of synthetic vowels that would be acceptable to native speakers of the language. By way of preparation for this and for clarification of problems encountered in the first attempts at synthesis, Thai subjects were asked to identify and comment on isolated vowels of real speech as well as vowels that had been passed through the Vocoder to produce one set on a monotone and another set with the original pitch regenerated. Then two-formant vowels were synthesized on the

Pattern Playback and tested perceptually until optimum formant frequencies were achieved.

1. Real Speech

Although phonemic analysis clearly establishes nine vowel phonemes, before Thai listeners could be asked to identify isolated synthetic vowels, it was necessary to test the assumption that they could identify the vowel utterances of a given speaker.

Test 1

Eighteen vowel utterances, nine single vowels on the low tone and nine geminate ones on the mid tone, were recorded on tape by W. N.[8] These stimuli were arranged in random order and presented to nine Thai subjects, who were told to write each spoken vowel in Thai script. A confusion matrix of the results is given in Figure 1.16. One person was responsible for the two confusions of the high central and mid central vowels. The results give experimental confirmation of the empirical observation that Thai speakers readily identify citation forms of vowels.

The Vocoder (Introduction C-1) was used as a link between tests on the perception of real speech and the perception of synthetic speech.[9] Its purpose was to help clear up difficulties met with in early attempts at synthesis.

Test 2

W. N.'s eighteen vowel utterances were put through the Vocoder, which was set to follow the pitch of his voice in order to keep the distinction between the mid tone of the geminate vowels and the low tone of the single ones. Recordings of the Vocoder output were arranged in random order, and the nine subjects were again asked to write the vowels that they heard. As in Test 1, the response categories were the nine single vowels on the low tone and the nine geminate vowels on the mid tone.

The results, shown in Figure 1.17, indicate that highly intelligible isolated Thai vowels can be had from the kind of analysis and synthesis performed by the Vocoder. That there was an 11% drop in intelligibility from Test 1 to Test 2 is not surprising, since the best of Vocoders lose some of the spectral characteristics of the speech input. Qualitative confusions are found in the matrix only between vowels separated by one distinctive feature of height (Introduction B-2a).

TEST I

VOWELS SPOKEN

VOWELS HEARD

	i	ii	e	ee	æ	ææ	ɨ	ɨɨ	ə	əə	a	aa	u	uu	o	oo	ɔ	ɔɔ	Totals
i	9	1	1																11
ii		8																	8
e			8																8
ee				9															9
æ					9														9
ææ						9													9
ɨ							9		1										10
ɨɨ								9		1									10
ə									8										8
əə										8									8
a											9								9
aa												9							9
u													9						9
uu														9					9
o															9				9
oo																9			9
ɔ																	9		9
ɔɔ																		9	9
Totals	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	162

Heard as spoken: 97.5%

FIG.1.16 CONFUSION MATRIX OF VOWELS AS PERCEIVED BY 9 SUBJECTS IN TEST I.

TEST 2

VOWELS SPOKEN

VOWELS HEARD

	i	ii	e	ee	æ	ææ	ɨ	ɨɨ	ə	əə	a	aa	u	uu	o	oo	ɔ	ɔɔ	Totals
i	7	3	1																11
ii		6																	6
e	2		6	3															11
ee			2	6		1													9
æ					9	1													10
ææ						7													7
ɨ							9												9
ɨɨ								8											8
ə									9										9
əə								1		9									10
a											9								9
aa												9							9
u													9						9
uu														5					5
o															9				9
oo														4		9			13
ɔ																	9	4	13
ɔɔ																		5	5
Totals	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	162

Heard as spoken: 86.4%

FIG.1.17 CONFUSION MATRIX OF VOCODED VOWELS WITH PITCH RETAINED AS PERCEIVED BY 9 SS IN TEST 2.

The striking length confusions in a few of the vowels were probably mainly due to the nature of the buzz-hiss switching, an instrumental problem of long standing in Vocoder operation. If the intensity of a voiced input is too low, the buzz generator simply does not respond, and the hiss is turned on instead. The vowels /ii ee ɔɔ/, and perhaps also /ææ/, must have given an impression of shortness to some of the listeners through a weakening of voicing toward the end with a resultant loss of buzz. Indeed, the productions of /ii/ and /ɔɔ/ used in Tests 1-4 were unusually short for isolated geminate vowels to begin with, although they caused no trouble in Test 1 (except for one identification of /ii/as /i/). The two identifications of /e/ as /ee/ in Test 2 are harder to explain. In any event, these results served as a control for the next two tests, which were to help explain the very large number of length confusions in the first of the Pattern Playback tests (Sec. 2, Test 5).

Test 3

This test was the same as Test 2 except that the Vocoder buzz was kept at a monotone of 120 cps. The intention was to simulate some of the conditions of a preliminary Pattern Playback test (to be described), in which the directions specified a tonal difference between the single and geminate vowels, although both were on a monotone.

After only four subjects had taken this test, it seemed quite clear (Fig. 1.18) that the absence of an expected pitch difference was causing a considerable increase in the number of length confusions. Clarification of this observation was sought in the next test.

Test 4

This test was simply a repetition of Test 3, except that the nine subjects were told that all the vowels were on the low tone.

The matrix in Figure 1.19 reveals a return in intelligibility to the level of Test 2. Of the six length errors remaining, only two were made by subjects who had taken Test 3. The percentages of length errors for all four tests (including the real-speech case, Test 1), are given in Table 1.5.

TABLE 1.5

Percentages of Length Errors in Tests 1-4

Test 1	Test 2	Test 3	Test 4
0.6%	8.0%	26.3%	3.7%

TEST 3

VOWELS SPOKEN

VOWELS HEARD

	i	ii	e	ee	æ	ææ	ɨ	ɨɨ	ə	əə	a	aa	u	uu	o	oo	ɔ	ɔɔ	Totals
i	3	3																	6
ii		1																	1
e	1		3	2															6
ee			1	2															3
æ					4	2													6
ææ						2													2
ɨ							4												4
ɨɨ								4											4
ə									4	2									6
əə										2									2
a											4	3							7
aa												1							1
u													4	2					6
uu														1					1
o															4				4
oo														1		4			5
ɔ																	4	4	8
ɔɔ																			0
Totals	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	72

Heard as spoken: 70.8%

FIG. 1.18 CONFUSION MATRIX OF VOCODED VOWELS ON A MONOTONE AS PERCEIVED BY 4 SS IN TEST 3.

TEST 4

VOWELS SPOKEN

VOWELS HEARD

	i	ii	e	ee	æ	ææ	ɨ	ɨɨ	ə	əə	a	aa	u	uu	o	oo	ɔ	ɔɔ	Totals
i	8																		8
ii		7																	7
e	1		9	1															11
ee		2		8															10
æ					8	1													9
ææ						8													8
ɨ							8												8
ɨɨ								6											6
ə					1		1		9										11
əə								3		8									11
a											9								9
aa										1		9							10
u													7						7
uu														6					6
o													2		9				11
oo														3		7			10
ɔ																1	9	3	13
ɔɔ																1		6	7
Totals	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	162

Heard as spoken: 87.0%

FIG. 1.19 CONFUSION MATRIX OF VOCODED VOWELS ON A MONOTONE AS PERCEIVED BY 9 SS IN TEST 4 WITH /ʌ/ SPECIFIED IN THE INSTRUCTIONS.

The only single vowel confused with a geminate in these tests was /e/. On the other hand, all the geminate vowels but /ɨɨ/ suffered confusion with single vowels in one test or another.

The results of the Vocoder tests made it evident that in experiments on the perception of Thai vowel qualities in which phonemic length plays a part, the tones specified in instructions to subjects should be appropriate to the pitch of the synthetic carrier. That this should have been obvious may be so, yet it is felt that the details are not without perceptual and linguistic interest. Vowel duration is the subject of Chapter II and will not be discussed further here.

Qualitative confusions in the foregoing tests ranged from 1.9% to 9.3%, leaving little room for interpretation except to say that the Vocoder effected a small loss in qualitative distinctions. Except for one confusion of /æ/ with /ə/ in Test 4, all qualitative errors involved pairs of vowels separated only by one distinctive feature of height. There was only one double error (length and quality); /oo/ was heard as /ɔ/ once in Test 4.

2. Artificial Speech

It has been found that two formants are enough to synthesize vowels quite well.[10] At this point experimentation was begun to obtain optimum formant frequency values for a set of acceptable two-formant synthetic Thai vowels. The speech synthesizer used for this purpose was the Pattern Playback (Introduction C-1).

The first set of vowels painted was based on the average frequency values of the first and second formants of W. N.'s vowels, the only full set of measurements available at the time (Fig. 1.10). These formant frequencies rounded to the nearest 5 cps, along with the closest possible Pattern Playback approximations, are given in Table 1.6.

TABLE 1.6

	W. N.		Pattern Playback	
	F-1 (cps)	F-2 (cps)	F-1 (cps)	F-2 (cps)
i	365	2100	360	2100
ii	300	2225	300	2220

TABLE 1.6 — Continued

	W. N.		Pattern Playback	
	F-1 (cps)	F-2 (cps)	F-1 (cps)	F-2 (cps)
e	525	1950	540	1980
ee	465	1975	480	1980
æ	700	1750	720	1740
ææ	675	1800	660	1800
ɨ	445	1250	420	1260
ɨɨ	445	1350	420	1380
ə	535	1175	540	1200
əə	540	1250	540	1260
a	720	1225	720	1200
aa	720	1175	720	1200
u	355	695	360	720
uu	375	720	350	720
o	480	830	480	840
oo	480	830	480	840
ɔ	620	980	600	960
ɔɔ	580	950	600	960

The eighteen vowels were painted at the formant frequencies listed in the right-hand columns of Table 1.6. Each formant had the width of two Pattern Playback channels. W. N.'s average vowel durations of 140 msec for single vowels and 475 msec for geminate vowels were used. Sample painted vowels are shown in Figure 1.20.

Test 5

Two paired presentations of each of the eighteen synthetic vowels were arranged in random order on a tape, making thirty-six test items. Seven subjects were instructed to identify each pair as a Thai vowel and write it in Thai script. Orthographic forms that indicate the mid tone for geminate vowels and the low

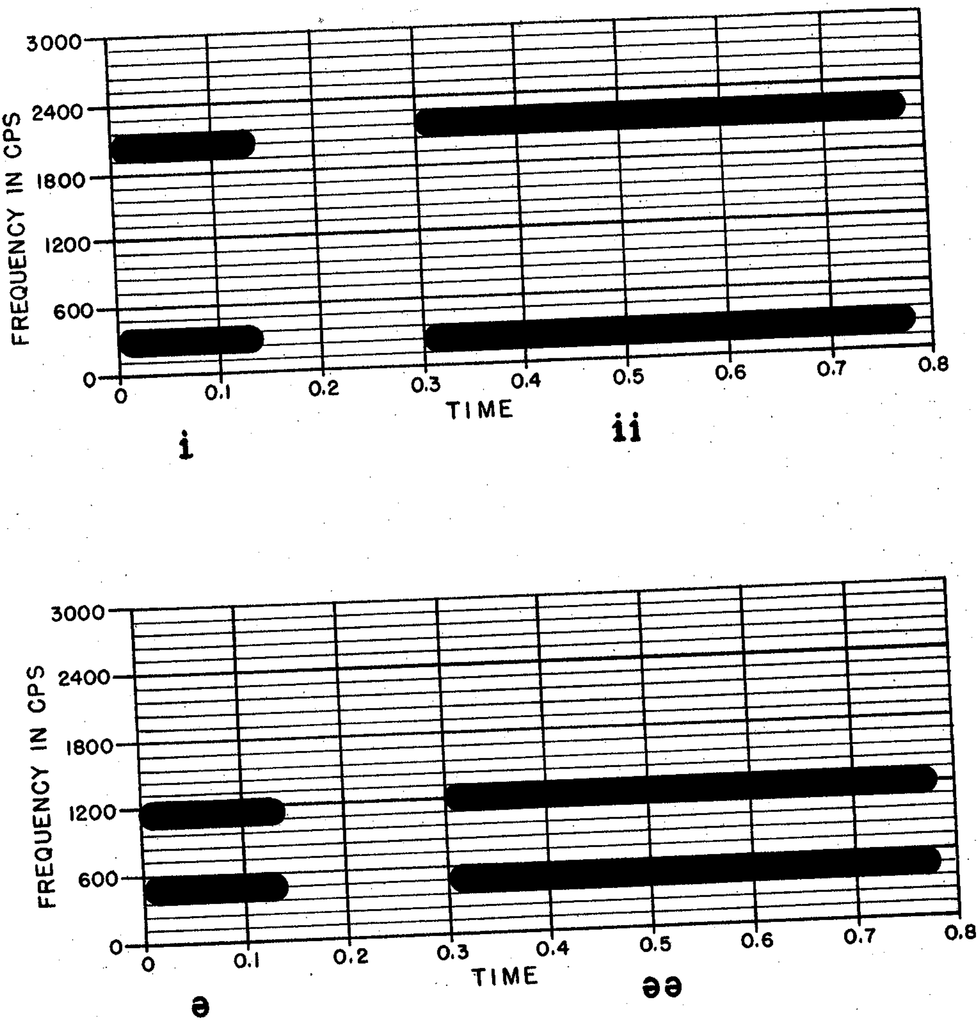


FIG. 1.20 EXAMPLES OF SYNTHETIC VOWEL FORMANTS IN TEST 5

TEST 5

VOWELS INTENDED

VOWELS HEARD

	i	ii	e	ee	æ	ææ	ɨ	ɨɨ	ə	əə	a	aa	u	uu	o	oo	ɔ	ɔɔ	Totals
i	6	3	2		1								2			1			15
ii		4										1						1	6
e			4		2				2		1		1		3		3		16
ee				5		1				1				1				2	10
æ			2		4														6
ææ																			0
ɨ	1		2		1	1	3	2	3	3		1	3	1	2		1		24
ɨɨ		2		3		2		2	1	1			1	3		2	2		19
ə			2		1			1	1		1	1	1		1		1		10
əə		1		1		2		1		2		1						4	12
a	3		1					1			2								7
aa		1		2				1				1							5
u	2	1			3		10		6		1		3	3	5	2	2		38
uu		2				2				2				3		5	1		15
o	1		1		1	1				1	7	1	1	1		2	4	5	26
oo	1			1	1	2		3		2		4	1	1	2	2		1	21
ɔ						1		1	1	1	1	2							7
ɔɔ				2		1		1			1			1	1				7
Totals	14	14	14	14	14	13	13	13	14	13	14	12	13	14	14	14	14	13	244

Heard as intended: 17.2 %

FIG. 1.21 CONFUSION MATRIX OF PATTERN PLAYBACK VOWELS IN TEST 5 WITH 7 SS.

tone for single vowels were heedlessly listed as the inventory of possible answers, because Thai vowels are normally named that way.[11]

A confusion matrix of the responses is given in Figure 1.21. This chaotic display, including confusions of both length and quality, seemed too much to explain merely on the basis of unfamiliarity with synthetic speech, although this was surely part of it. It was at this point that recourse was had to the Vocoder (Sec. 1).

Test 6

The purpose of this test was to get impressionistic reactions to the stimuli of Test 5 to complement the findings of the Vocoder experiments. E. N. was asked to listen to each stimulus as many times as he wished, identify it and comment on it. It did not take him long to point out that the sameness of pitch from one stimulus to another so contradicted the tonal distinction present in the instructions as to confuse him considerably. Another disturbing factor was the abruptness with which both short and long vowels ended; that is, the abrupt terminations of the painted formants suggested the presence of a glottal stop not only at the ends of single vowels, where it normally occurs, but also at the ends of geminate ones, where it does not occur.[12] E. N. was quite convinced that the numerous length confusions and probably much of the general confusion in Test 5 could be blamed on these two factors.

Once these matters had been clarified, E. N. was able to identify most of the synthetic vowels very well and to make helpful comments on vowel qualities. There were, nevertheless, a few vowels that gave him a great deal of trouble, even after many repetitions. These vowels were also troublesome in Test 7 and the formant frequencies were changed for Test 8.

Test 7

This test was simply a new presentation of Test 5 with a change in the instructions but no change in the Pattern Playback stimuli. This time ten subjects, including six who had taken Test 5 a month and a half before, but excluding E. N., were told that all the vowels were on the low tone.

The results are given in the confusion matrix in Figure 1.22. The improvement in overall intelligibility from 17.2% to 69.9% is startling. The four new subjects, for whom the sounds must have been as strange as they had been for the other six people in Test 5, were responsible for only slightly more than their proportionate share of the errors. Although length discrimination was improved, there still were several clusterings of errors, with geminate vowels being taken for their single counterparts or some other single vowel.

TEST 7

VOWELS INTENDED

VOWELS HEARD

	i	ii	e	ee	æ	ææ	ɨ	ɨɨ	ə	əə	a	aa	u	uu	o	oo	ɔ	ɔɔ	Totals
i	20							1											21
ii		20		5															25
e			20	3	6	3													32
ee				12		10													22
æ					14														14
ææ						3													3
ɨ						1	12	1											14
ɨɨ								8		2									10
ə						2	7	3	20	4		3							36
əə								6		14									23
a											15								15
aa								1				12							13
u													19		2	1			22
uu														13		2			15
o													1	1	18	1	6		27
oo														6		14		17	37
ɔ						1					5	1					14		21
ɔɔ												4				2		3	9
Totals	20	20	20	20	20	20	19	20	20	20	20	20	20	20	20	20	20	20	359

Heard as intended: 69.9%

FIG. 1.22 CONFUSION MATRIX OF PATTERN PLAYBACK VOWELS IN TEST 7 WITH 10 SS.

Test 8

Two kinds of changes were made in the stimuli in this experiment: (1) Tapered endings were given to the formants of all geminate vowels to remove the impression of a glottal stop reported in Test 6. Single vowels retained the abrupt endings shown in Figure 1.20. (2) New formant frequencies were given to the vowels that E. N. had trouble in identifying in Test 6. The frequencies were chosen on the basis of his careful criticism of the vowel qualities produced in comparison with those intended.[13] The vowels thus adjusted are listed in Table 1.7 with their original formant frequencies and the new ones.

TABLE 1.7

	Tests 5-7		Test 8	
	F-1 (cps)	F-2 (cps)	F-1 (cps)	F-2 (cps)
ææ	660	1800	720	1800
aa	720	1200	780	1260
uu	360	720	300	660
ɔɔ	600	960	660	960

Not only these synthetic vowels, but others that E. N. readily identified, caused trouble in Test 7. The latter, /ee æ ɨ ɨɨ a oo ɔ/, were left unadjusted at this stage.

The results of Test 8 are shown in Figure 1.23. Length discrimination is better, but not much better, than that of Test 7. Percentages of stimuli confused as to length in the three tests are given in Table 1.8.

TABLE 1.8

Percentages of Length Errors in the Synthetic Vowel Tests

Test 5	Test 7	Test 8
20.2%	6.5%	4.2%

TEST 8

VOWELS INTENDED

VOWELS HEARD

	i	ii	e	ee	æ	ææ	ɨ	ɨɨ	ə	əə	a	aa	u	uu	o	oo	ɔ	ɔɔ	Totals
i	13	1	1																15
ii		12																	12
e			11	1	4														16
ee				12															12
æ					9	2													11
ææ						12													12
ɨ		1	1				7	1						1					11
ɨɨ								5		1									6
ə			1		1		7	2	14	1									26
əə								6		12									19
a											13								13
aa												13						2	15
u	1												13	1	3				18
uu														8					8
o													1		11	1	5		18
oo														1		12			13
ɔ											1						9		10
ɔɔ												1		3		1		12	17
Totals	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	252

Heard as intended: 79.0 %

FIG. 1.23 CONFUSION MATRIX OF PATTERN PLAYBACK VOWELS IN TEST 8 WITH 7 SS.

It would seem that changing the specification of tones was more important than changing the formant endings in the geminate vowels, although, in the absence of a control group that had the latter change before the former, the evidence is not conclusive. In any event, E. N. reported that he felt less strain with the new stimuli in distinguishing long vowels from short ones.

In general, the changes in the formant frequencies of the vowels in Table 1.7 effected an improvement in identifiability, but the response to /uu/ was still bad. The responses to /ee a oo/, not too satisfactory in Test 7, were good in Test 8. /æ ɨ ɨɨ ɔ/, however, continued to give difficulty.

Test 9

This test was run to evaluate replacements for the five defective synthetic vowels. It was given in two parts: (1) identification of stimuli, as before, and (2) evaluation of the stimuli presented in (1).

Examination of the three confusion matrices (Figs. 1.21-23) and consideration of E. N.'s criticism of the vowel qualities in Test 6 led to the choice of formant frequency values in Table 1.9. They include previously tried values as well as new ones. The second formant for two variants of synthetic /uu/ was reduced in intensity by repainting with narrower lines (Introduction C-1).

TABLE 1.9

	F-1 (cps)	F-2 (cps)	Choice
æ	720	1740	
	720	1800	†
	780	1800	*
ɨ	420	1260	
	300	1260	†
	300	1380	*
	300	1500	†
ɨɨ	480	1380	
	420	1380	
	300	1380	*
	300	1260	†
	300	1500	†

TABLE 1.9 — Continued

	F-1 (cps)	F-2 (cps)	Choice
uu	300	660	
	300	660 (weak)	*
	360	720	
	360	720 (weak)	
ɔ	600	960	†
	660	960	*

* Final preference
† Also good

In the first part of Test 9, two paired occurrences of each of the eighteen stimuli, making thirty-six test stimuli, were presented in random order to seven subjects, who took the test one at a time. For the second part of the test, given immediately thereafter, all eighteen stimuli were available on bits of magnetic tape affixed to cards for the Card Reader (Introduction C-1, Tape recorders). If the subject had identified more than one stimulus as a particular vowel in the first half of the test, he was told to play the cards bearing these recorded stimuli as many times as he wished and choose the best production of that vowel. The final preferences, expressed in every instance by at least six of the seven subjects, are marked with asterisks in the right-hand column of Table 1.9. Daggers indicate productions considered nearly as good or as good by at least four subjects.

The formant frequencies of the final set of two-formant synthetic Thai vowels that emerged from the foregoing experiments are shown in Table 1.10.

TABLE 1.10

Synthetic Thai Vowels

	F-1 (cps)	F-2 (cps)		F-1 (cps)	F-2 (cps)
i	360	2100	ii	300	2220

TABLE 1.10 — Continued

	F-1 (cps)	F-2 (cps)		F-1 (cps)	F-2 (cps)
e	540	1980	ee	480	1980
æ	780	1800	ææ	720	1800
ɨ	300	1380	ɨɨ	300	1380
ə	540	1200	əə	540	1260
a	720	1200	aa	780	1260
u	360	720	uu	300	660
o	480	840	oo	480	840
ɔ	660	960	ɔɔ	660	960

There is a logarithmic plot of the synthetic vowels in Figure 1.24.

C. Discussion of Results

Acoustic analysis has yielded data on the formant frequencies of the vowels of Standard Thai. This study has concentrated on citation forms of vowels, which may be said to be norms in Thai speech. It is hoped that these data will serve as a point of departure for a supplementary investigation of vowels in running discourse. A start in this direction has been made in Section A-3 of this chapter.

A striking feature of these citation forms is the prevalence of steady-state vowels, which can be seen in the spectrograms of Figures 1.2-8. Of the two speakers' utterances examined, 75% were steady states. Of the citation forms with shifting formants, 80% were geminate vowels. Of the latter, 70% were on dynamic tones.

Examination of the formant frequencies of the final portions of the citation forms of the vocalic clusters /ia ɨa ua/ (Table 1.3) shows that the formant-frequency range of these manifestations of /a/ tends to approach that of /ə/. Indeed, in connected speech these clusters often end in phones that are similar to allophones of /ə e o/ and perhaps other vowels. Some examples appear in Table 1.4. Two possible explanations for this that have not been

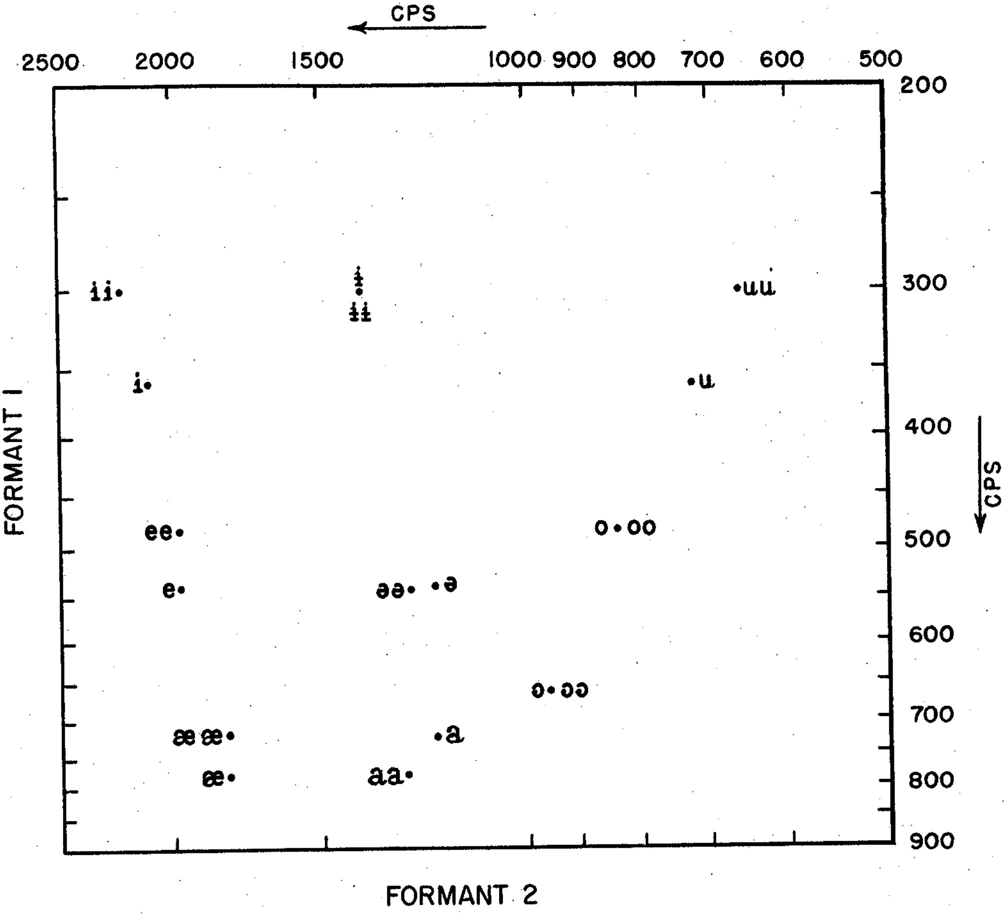


FIG. 1.24 SYNTHETIC THAI VOWELS

investigated in this study may be advanced: (1) The final element in vocalic clusters is a reduced vowel representing a neutralization of vocalic features. (2) In these clusters, /a/ and other vowels are partly in free alternation and partly in phonologically conditioned alternation, in which case the transcriptions /ia ɨa ua/ would be morphophonemic spellings.

As expected, the sampled connected speech shows a wider range of formant frequencies and somewhat less neat separation of the vowels in the acoustic space than do the citation forms; nevertheless, the systematic relations are the same. An extension of this study to obtain formant frequencies in each type of phonemic environment for comparison with the citation forms would be of some interest.

Speech synthesis was used to establish appropriate frequencies of the first and second formants of vowels that are readily identified as the equivalents of the spoken Thai vowels and to provide a comparison between synthetic two-formant vowels and spoken multiformant vowels. The formant frequencies of the synthetic vowels are essentially independent values that were obtained through perception testing; they were changed in accord with test results until optimum values were achieved. These final values show rather good agreement with real speech, as can be seen in Table 1.11, which displays the synthetic formant frequencies in direct comparison with the average values of the first and second formants of the vowels spoken by W. N. and E. N. That there is closer agreement with W. N.'s values may not be too surprising, since his formant frequencies served as a point of departure for the synthesis.

TABLE 1.11

	Synthesis		W. N.'s Av.		E. N.'s Av.	
	Single Vowels					
	F-1 (cps)	F-2 (cps)	F-1 (cps)	F-2 (cps)	F-1 (cps)	F-2 (cps)
i	360	2100	365	2100	365	2140
e	540	1980	525	1950	415	1860
æ	780	1800	700	1750	590	1640
ɨ	300	1380	445	1250	360	1360
ə	540	1200	535	1170	525	1240
a	720	1200	720	1225	660	1325

TABLE 1.11 — Continued

	Synthesis		W. N.'s Av.		E. N.'s Av.	
	Single Vowels					
	F-1 (cps)	F-2 (cps)	F-1 (cps)	F-2 (cps)	F-1 (cps)	F-2 (cps)
u	360	720	355	695	355	755
o	480	840	480	830	515	885
ɔ	660	960	620	980	620	990
	Geminate Vowels					
ii	300	2220	300	2225	310	2240
ee	480	1980	465	1975	420	2000
ææ	720	1800	675	1800	625	1890
ɨɨ	300	1380	445	1350	375	1450
əə	540	1260	540	1250	510	1310
aa	780	1260	720	1175	670	1420
uu	300	660	375	720	350	720
oo	480	840	480	830	470	910
ɔɔ	660	960	580	950	610	970

This information can be viewed graphically by comparing Figures 1.12-13 with Figure 1:24.

Notes

1. A slight discrepancy that can have no bearing on the outcome must be noted. At the outset of this investigation, I was misled by considerations of lexical frequency into thinking that a single vowel on the high tone was an abnormal utterance, and so I had W. N. record single vowels as /V̀ hV̀ hV́/. Since /h/ is nothing but a kind of weak-energy, voiceless extension of the formants, the resulting vowels are essentially the same as completely isolated ones. My second informant, E. N., recorded only /V̀ V́/.

2. For a discussion of the importance of sustained vowels in speech study, see G. E. Peterson, "The Phonetic Value of Vowels," Lg., XXVII (1951), 541-553.

3. Pierre Delattre, "Les indices acoustiques de la parole: Premier rapport," Phonetica, II (1958), 245.

4. See any statistics book for a discussion of this, e.g., M. J. Moroney, Facts from Figures (Harmondsworth, Middlesex, 1951), chap. 5.

5. For earlier uses of such a two-dimensional display, see Hugo Pipping, Über die Theorie der Vocale (Helsinki, 1895), p. 29, and Inledning till Studiet av de Nordiska Språkens Ljudlära (Helsinki, 1922), p. 37; A. W. de Groot, "Phonologie und Phonetik als Funktionswissenschaften," TCLP, IV (1931), 119-120; John Lotz, "Notes on Structural Analysis in Metrics," Helicon, IV (1942), 121; Martin Joos, Acoustic Phonetics (Baltimore, 1948), pp. 50-65.

6. Where a glide was plotted, as in /hà/, only the mid 50% of the vowel was considered in deriving the averages from the scatter plots.

7. For exhaustive tables, see M. Fowler and T. Israsena, The Total Distribution of the Sounds of Siamese (Madison, 1952).

8. The mid tone was chosen for the geminate vowels partly because long vowels are traditionally cited that way and partly because "tonally neutral" utterances seemed desirable for these experiments. Short vowels are traditionally cited on the low tone; in isolation they do not take the mid tone. This tonal dichotomy turned out to be an unwanted variable in later tests.

9. Vocoder experiments might just as well be discussed under "Artificial Speech" (Sec. 2). They are included here, because they involve a direct manipulation of a real-speech input.

10. P. Delattre, A. M. Liberman, F. S. Cooper and L. G. Gerstman, "An Experimental Study of the Acoustic Determinants of Vowel Color; Observations on One- and Two-Formant Vowels Synthesized from Spectrographic Patterns," Word, VIII (1952), 195-210.

11. Unless specific tonal marks are included, Thai script automatically specifies the low tone for single vowels and the mid tone for double vowels; the Pattern Playback, however, produced both lengths on the same pitch.

12. Except with the high and falling tones, but they were not used in the tests.

13. Without any previous phonetic training, E. N. turned out to be a highly perceptive informant. His keen observation and interest were helpful throughout the study.

II. VOWEL DURATION

A. Measurements

In the phonology outlined in the Introduction, vowel length in Thai has been analyzed as a sequence of two vowels versus a single vowel. It is posited that on the phonetic level the relevant cue is duration, although other allophonic features are concomitant with occurrences of both double vowels (geminates and vocalic clusters) and single vowels.

In this chapter an investigation of the quantitative distinction between single and geminate vowels is described. Vocalic clusters are brought in for comparison with geminates. Another theory on vowel length is briefly held up for criticism in the Discussion of Results.

1. Citation Forms of Vowels

The sets of spectrograms of citation forms of Thai vowels (see Figs. 1.2-8) measured for formant frequencies in Chapter I were also measured for vowel durations. Duration was measured as accurately as possible in each vowel utterance. For a discussion of precision in this kind of measurement, see I. A-1.

The durations of vowels spoken by W. N. and E. N. specified to the nearest 5 milliseconds, are set forth in Table 2.1, which corresponds in arrangement of entries to Tables 1.1 and 1.2. The symbol /h/ with single vowels applies only to W. N. (I. f.n. 1).

TABLE 2.1

Durations of Citation Forms of Vowels in Msec

	W. N.	E. N.
ì	120	130
(h) ì	140	
(h) í	90	140

TABLE 2.1 — Continued

	W. N.	E. N.
ii	410	390
ìi	530	380
íi	440	345
îi	440	340
ǐi	450	340
è	150	140
(h) è	150	
(h) é	80	160
ee	420	400
eè	500	360
eé	410	320
eê	385	340
eě	450	340
æ	140	150
(h) æ̀	150	
(h) ǽ	100	150
æ æ	500	420
æ̀æ	530	400
ǽæ	440	350
æ̂æ	430	380
æ̌æ	490	380
ɨ̀	185	180
(h) ɨ̀	165	
(h) ɨ́	100	160
ɨɨ	520	400
ɨ̀ɨ	520	400
ɨ́ɨ	470	340
ɨ̂ɨ	440	340
ɨ̌ɨ	480	400
ə̀	190	150
(h) ə̀	130	
(h) ə́	100	155
əə	550	360
əə̀	560	440
əə́	420	340
ə̂ə	470	320
əə̌	510	400

TABLE 2.1 — Continued

	W. N.	E. N.
à	195	190
(h) à	190	
(h) á	80	240
aa	540	460
àa	560	480
aá	440	400
aâ	505	340
aǎ	455	420
ù	170	140
(h) ù	170	
(h) ú	100	150
uu	480	390
ùu	540	420
uú	465	360
uû	510	380
uǔ	520	340
ò	200	160
(h) ò	160	
(h) ó	130	160
oo	500	380
òo	500	420
oó	440	340
oô	460	330
oǒ	500	360
ɔ̀	180	180
(h) ɔ̀	140	
(h) ɔ́	95	170
ɔɔ	510	440
ɔ̀ɔ	480	420
ɔɔ́	400	370
ɔɔ̂	445	320
ɔɔ̌	450	360
Averages		
V	140	160
VV	475	380
VV/V ratio	3. 39	2. 38

The distribution of these data is displayed graphically in Figure 2.1 A, B, with the durations rounded to the nearest 10 msec.

2. A Sampling of Vowels in Connected Speech

a. Vocalic clusters

Since long vowels were analyzed as geminates by analogy with vocalic clusters, it should be of some interest to have data on the durations of the latter for ready comparison with the former. The citation forms of the vocalic clusters that furnished formant frequencies for Table 1.3 were measured for durations. The durations of these clusters, spoken on all five tones by E. N., are presented in Table 2.2.

TABLE 2.2

Durations of E. N.'s Citation Forms of Vocalic Clusters in Msec

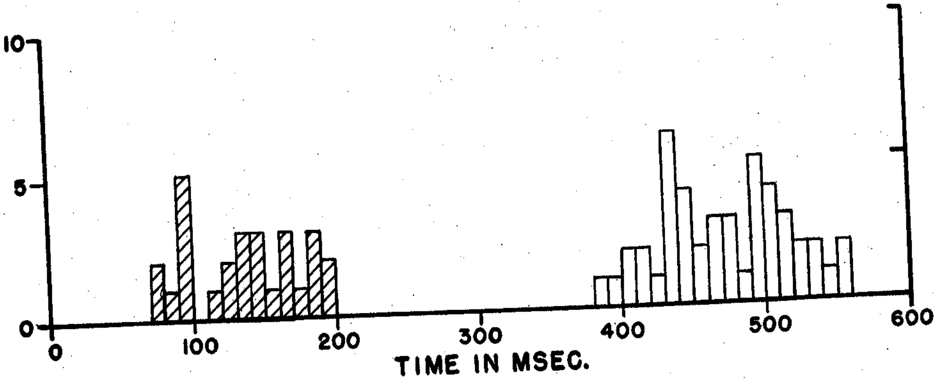
ia	500	ɨa	500	ua	460
ìa	520	ɨ̀a	500	ùa	440
ía	400	ɨ́a	380	úa	340
îa	360	ɨ̂a	460	ûa	440
ǐa	440	ɨ̌a	500	ǔa	440

Average: 445

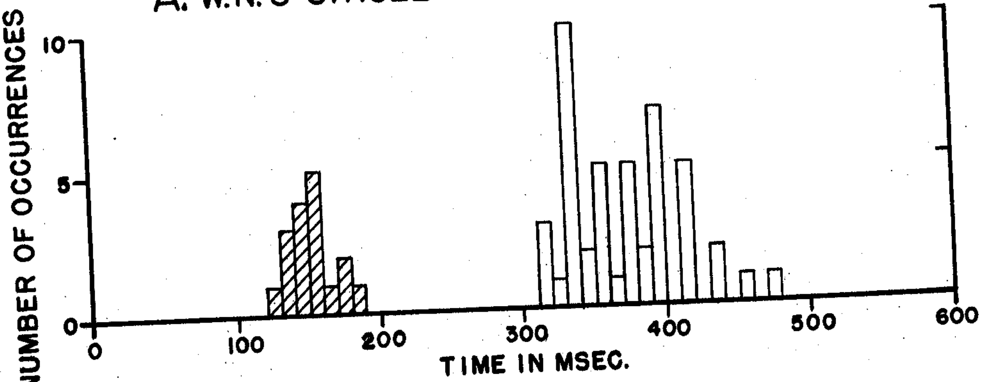
These data are displayed graphically in Figure 2.1C. Of E. N.'s double vowels, then, the average duration of the vocalic clusters is 14.6% longer than that of the geminates. Dissolving the clusters into component vowels for comparison with single vowels would be too difficult and arbitrary for any validity (see Fig. 1.14). The ratio of the clusters to E. N.'s single vowels (Table 2.1) is 2.78.

b. Vowels in running discourse

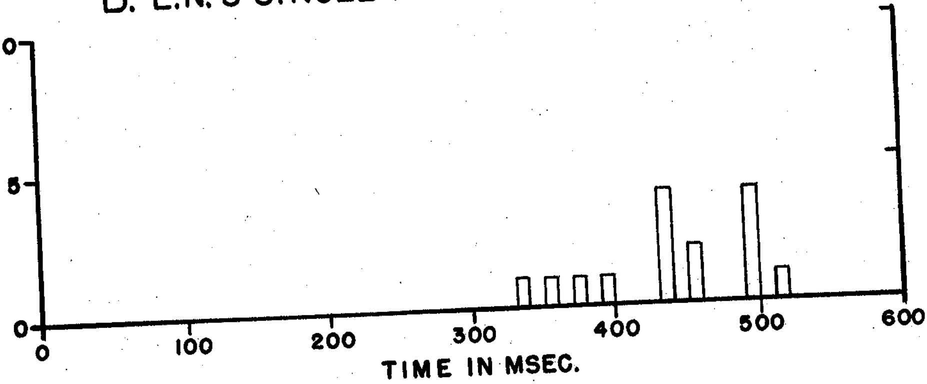
A thorough study of vowels in running discourse would require that for each set of measurements only vowels occurring under the same environmental and rhythmic conditions be considered. The measurements given here are not intended as a study of this kind;



A. W.N.'S SINGLE AND GEMINATE VOWELS



B. E.N.'S SINGLE AND GEMINATE VOWELS



C. E.N.'S VOCALIC CLUSTERS

FIG. 2.1 DURATIONS OF CITATION FORMS OF VOWELS rather, the data are simply meant to give some notion of the range of vowel durations in a variety of phonemic environments.

The narrative material described in I. A-4b furnished a sampling of vowel durations in connected speech. Table 2.3 contains the durations of the vocalic nuclei that were presented in terms of formant frequencies in Table 1.4. Here, as in Chapter I (A-4b), semivowels and consonant-vowel transitions are regarded as extensions of vowel formants and included in the measurements (see Fig. 1.15).

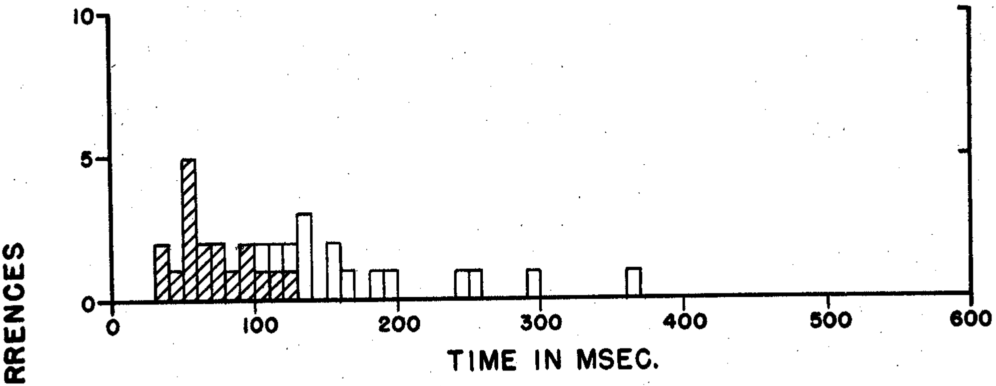
TABLE 2.3

E. N.'s Vowel Durations in Connected Speech

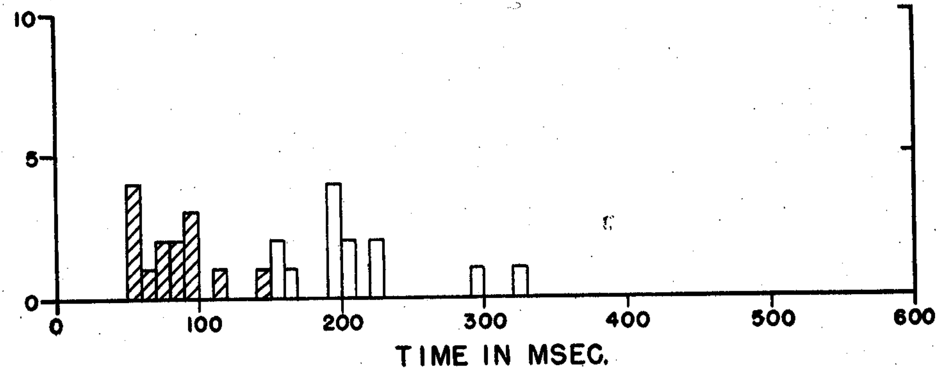
Vocalic Nucleus	Word	Duration in Msec
ji	dâjjin 'hear'	90
í	aatʿít 'week'	60
ii	dii 'to be good'	140
ií	nií 'this'	245
iǎ	siǎcaj 'to be sorry'	120
e	pen 'to be (something)'	55
é	lék 'to be small'	85
ee	ceedii 'stupa'	130
æ̀	pɨ̀kpʿæ̀n 'firmness'	100
æ̀	tæ̀ (∼tææ̀) 'but'	80
æǽw	læǽw 'finished'	210
ææ̂	mææ̂ 'mother'	160
ɨ̀	pɨ̀kpʿæ̀n 'firmness'	65
ɨ̌	tʿɨ̌ŋ 'to reach'	80
ɨ̌ɨ	nǎŋsɨ̌ɨ 'book'	300
ɨ̂a	mɨ̂a 'when'	370
ə	ŋən 'money'	110
əə	dəən 'to walk'	140

TABLE 2.3 — Continued

Vocalic Nucleus	Word	Duration in Msec
ə̀ə	bə̀ək ʻto withdraw (money)ʼ	140
əəj	kʻəəj ʻeverʼ	135
âw	kâw ʻnineʼ	160
a	samǎj ʻoccasionʼ	40
a	ma (~maa) ʻagoʼ	60
á	rák ʻto loveʼ	100
aa	naan ʻa long timeʼ	160
àa	ookàat ʻopportunityʼ	170
wáa	kʻónkʻwáa ʻto do researchʼ	250
âa	mâak ʻto be muchʼ	200
âa	baroo+mátʻâat ʻname of a monasteryʼ	190
ǎaj	lǎaj ʻseveralʼ	140
ù	sùt ʻmostʼ	60
jùu	jùu ʻto dwellʼ	200
ùa	kʻùap ʻyear (of a child's age)ʼ	125
o	kʻon ʻpersonʼ	50
o	okàat (~ookàat) ʻopportunityʼ	60
ó	kʻónkʼwáa ʻto do researchʼ	35
ǒ	pʻǒm ʻI (polite male)ʼ	70
ǒo	sukʻǒotʻaj ʻSukhothaiʼ	110
ɔ́	ŋɔ́ ʻnegritoʼ	120
ɔ̂ɔ	kɔ̂ɔ ʻthenʼ	260



A. VOWELS IN RUNNING DISCOURSE.



B. VOWELS IN MINIMAL PAIRS IN A CARRIER SENTENCE.

FIG. 2.2 DURATIONS OF VOWELS IN CONNECTED SPEECH

These data are displayed graphically in Figure 2.2A, with the durations rounded to the nearest 10 milliseconds. Nuclei with semivowels are excluded from the graphs. The shortest geminate occurring in the text is the /ǒo/ of /suk'ǒot'aj/, which is 110 msec long. This is the same as the /ə/ of /ŋən/ and shorter than the longest single vowel, the /ɔ́/ of /ŋɔ́/, which is 120 msec long, the same as the cluster /ǐa/ of /sǐacaj/. What with variations of speed and phonological conditioning, it is surprising that there was no more overlap in absolute durations than this in a sample drawn, without selection for length, from half an hour of colloquial speech.

3. Minimal Pairs in a Verbal Context

Partly to get a sampling of vowel-length contrasts in running speech under phonologically controlled conditions (Sec. 2-b) and partly to lay the groundwork for the perception tests in Part B of this chapter, a substitution frame was chosen which could serve as a carrier sentence for any minimal pair of Thai words. The frame was /p'ǒm p'ûut k'am wâa '_____ | hâj k'un faŋ↓/ 'I'm saying the word _____ for you to listen to.'[1]

Pairs of words minimally distinguished by vowel length could not readily be found for all nine vowels. Had it been possible to use obscene terms, some of the gaps might have been filled. The fourteen pairs of words that were embedded in the carrier sentence and recorded by E. N. included six of the vowel phonemes and all five tones. Spectrograms of one such pair of sentences can be seen in Figure 2. 3. The durations of the vowels are given in Table 2.4.

TABLE 2.4

Durations of Vowels in Minimal Pairs Spoken by E. N. in a Carrier Sentence

V	Word	Duration in Msec	VV	Word	Duration in Msec	Ratio VV/V
ì	cìp 'to taste'	60	ìi	cìip 'to fold (paper)'	170	2.83
ǐ	sǐn 'riches'	60	ǐi	sǐin 'moral code'	160	2.67
è	hèt 'mushroom'	90	èe	hèet 'reason'	210	2.33

TABLE 2.4 — Continued

V	Word	Duration in Msec	VV	Word	Duration in Msec	Ratio VV/V
ǽ	pʿǽ 'goat'	150	ǽæ	pʿǽæ 'to be defeated'	330	2.20
a	tʿan 'on time'	70	aa	tʿaan 'donation to charity'	200	2.86
a	baŋ 'to hide'	100	aa	baaŋ 'some'	220	2.20
à	bàt 'card'	100	àa	bàat 'Baht'	225	2.25
à	hàt 'to practice'	60	àa	hàat 'shoal'	200	3.33
á	pʿrá 'priest'	120	áa	pʿráa 'jungle knife'	300	2.50
á	kʿán 'to press'	90	áa	kʿáan 'to hinder'	200	2.22
â	kʿâm 'dusk'	80	âa	kʿâam 'to cross'	200	2.50
ǎ	kʿǎj 'to wind'	80	ǎa	kʿǎaj 'to sell'	160	2.00
ù	kʿùt 'to dig'	60	ùu	kʿùut 'to rub off'	210	3.50
ò	òt 'to abstain'	100	òo	òot 'to lament'	230	2.30
Averages:		87			215	2.47

In /kʿǎj/ and /kʿǎaj/ the full diphthongs were 260 msec long and 280 msec long respectively. It was possible to assign a segment to /j/ in each of these syllables and exclude it from the vowel measurements by finding a point where the formants diverged sharply from the steady-state (Fig. 2.4). This was done, contrary to earlier practice (Sec. 2b), because it was thought desirable to use at least one vocalic nucleus ending in a semivowel in the perception testing later, and taking into account only the overall durations of the two would have made the experiment meaningless. The durations of Table 2.4, rounded to the nearest 10 msec, are shown graphically in Figure 2.2B.

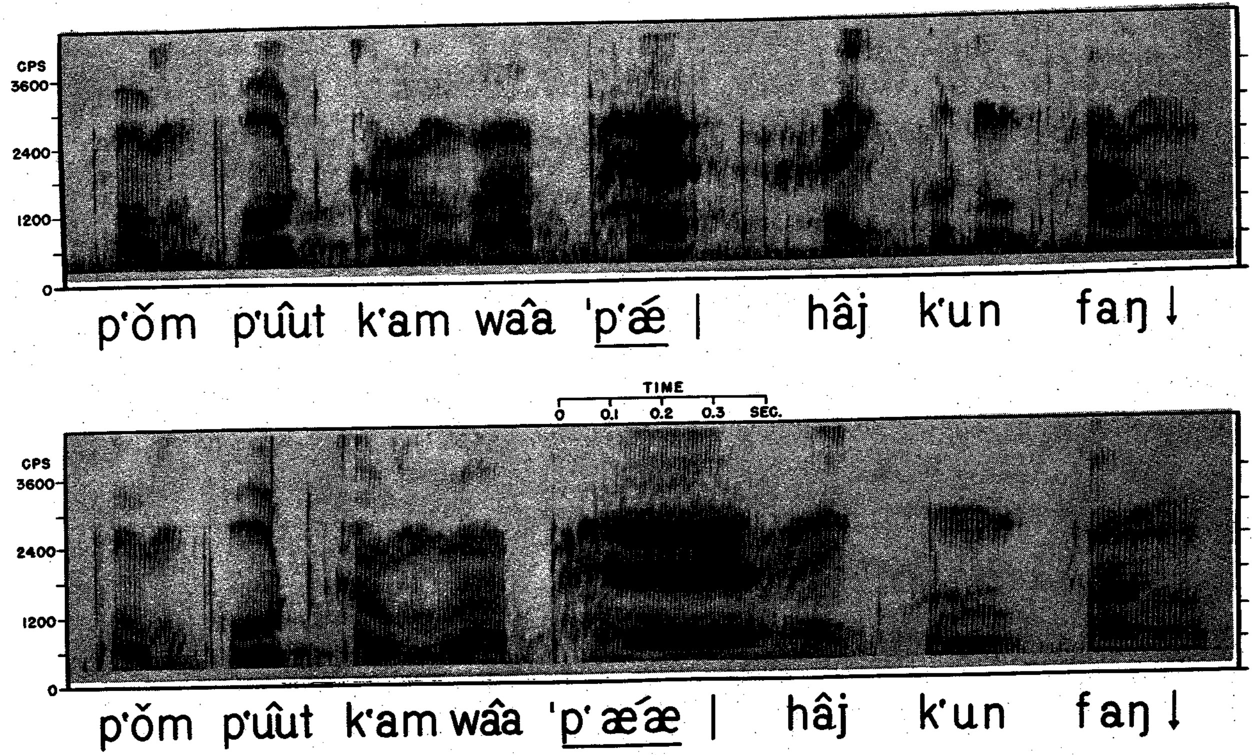


FIG. 2.3 SPECTROGRAMS OF /pʻǽ/ AND /pʻæǽ/ IN A CARRIER SENTENCE.

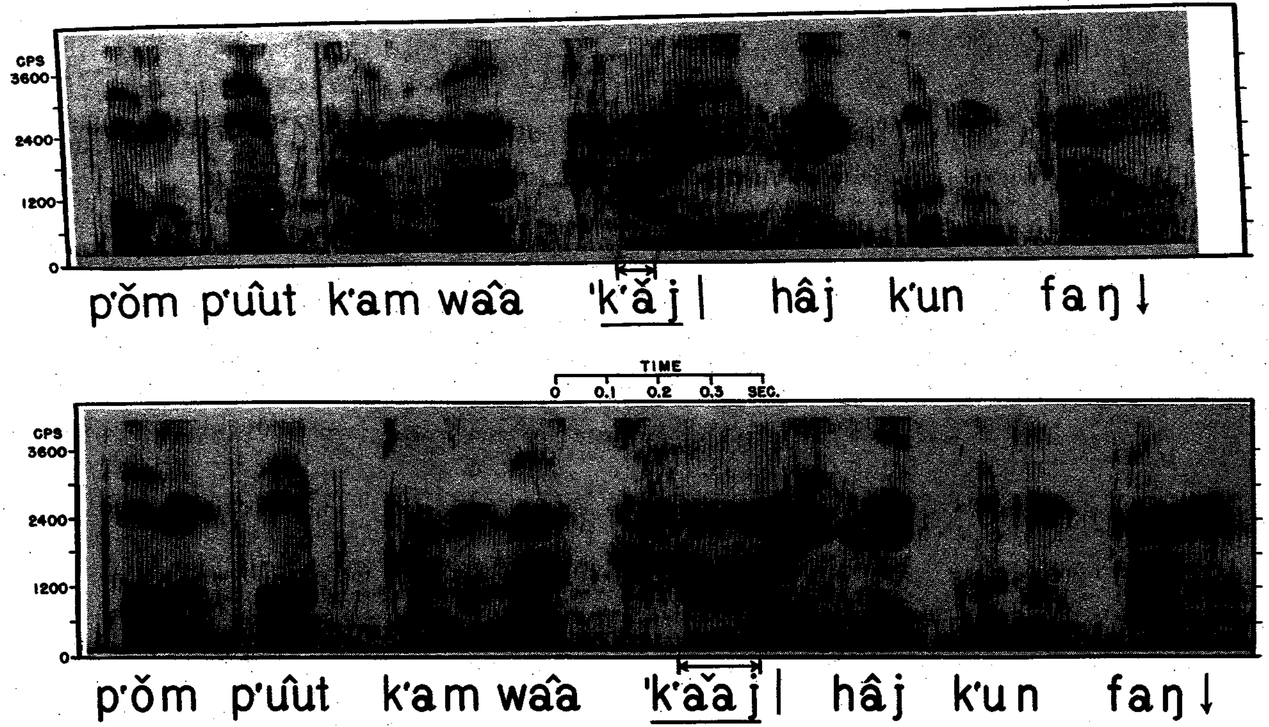


FIG. 2.4 SPECTROGRAMS OF /k'ǎj/ AND /k'aǎj/ IN A CARRIER SENTENCE WITH MEASURED DURATIONS INDICATED.

One of the pairs, /hàt/:/hàat/, was also recorded by K. C., a female informant. (/dìcʻǎn 'I (female)' was substituted for /pʻǒm/ 'I (male)' in the carrier sentence for her recording). Her /à/ was 130 msec long, and her /àa/, 360 msec, a ratio of 2.77.

In general, final semivowels and nasals in syllables containing single vowels were longer than those in syllables containing geminates. For example, the /n/ of /sǐn/ was 150 msec long, while the /n/ of /sǐin/ was 110 msec long. There is also the example of /kʻǎj/:/kʻǎaj/ in which the nearly equal durations of the vocalic nuclei are accounted for by the greater duration of the formant transitions for the /j/ of /kʻǎj/. There is usually enough discontinuity at one or more points between vowel formants and nasal resonances to distinguish them in spectrograms. The duration of /j/ was taken to be the duration of the remainder of each vocalic nucleus after the steady-state had been arbitrarily assigned to the vowel.

B. Experiments on the Perception of Quantitative Vowel Distinctions

If one generalizes from the measurements of the various sets of Thai vowels, double vowels in any new corpus are likely to be from 2 to 3.5 times as long as single vowels in analogous environments. Even in the sampling of vowels in running discourse (Table 2.3), where no attempt was made to control the phonemic environments of the contrast, the average ratio of double vowels to single vowels (excluding vowels preceded or followed by semivowels, as in Figure 2.3A), was 2.57. The impressionistic assignment of the distinction between single and geminate vowels to the feature of relative length seemed, therefore, to be well supported by the physical measurements.[2] It was nonetheless necessary to seek experimental verification of the relevance of relative length as the acoustic cue to the contrast.

1. Isolated Vowels

As a point of departure for the experiments in which length was to be the variable, a test was given to establish that all nine vowels were perceptually distinguishable as single vowels and geminates, as set forth in the phonemic analysis. This was done with citation forms of the vowels in connection with the experiments on the perception of vowel qualities described in Chapter I. The results of

Test 1 (Fig. 1.16) showed that Thai listeners had no difficulty in making this discrimination. In a total of 162 responses there was only one length confusion, one occurrence of /ii/ being taken for /i/.

The Vocoder and Pattern Playback stimuli employed in Tests 2-8 demonstrated the role of the glottal stop and the importance of the correct specification of tones in testing the discrimination of citation forms of single and geminate vowels.

2. Minimal Pairs

Both auditory impressions and acoustic measurements indicate a certain amount of variation in phonetic quality between single and geminate vowels.[3] It is also true that only three of the final set of two-formant synthetic vowels, /ɨ o ɔ/, had the same formant frequencies for both the single vowels and the geminates. Experiments were designed to test the hypothesis that even when such qualitative differences were left unaltered, listeners would hear vowels as single or geminate solely on the basis of length. A further goal was to determine perceptual crossover points dependent on duration between single and geminate vowels.

Tape cutting and splicing techniques (Introduction C-1) were used to remove successively longer central portions of the geminate vowel in each of the minimal pairs described in Section A-4 of this chapter. The general procedure was as follows: Many copies were made of the recording of each carrier sentence embedding a geminate vowel, and progressively longer segments, starting with 50 msec and increasing in steps of 25 msec, were cut from successive copies until the original geminate had been shortened to the length of the single vowel in the minimal pair. Where it was possible to do so without impairing the intelligibility of the word or the quality of the vowel, the shortening process was carried one or two steps below the duration of the single vowel. Naturally, the longer the geminate vowel, the greater the number of cuts that were made (see Table 2.4). The initial cut for each segment was made near the beginning of the vowel but far enough into it not to disturb any consonant-vowel transitions present. Cuts of eighty degrees were found most satisfactory for noise-free splices that did not extend far along the time axis. The overall precision of this method of shortening the vowels is approximately 1%. These shortened utterances, together with the original single and geminate vowels, were recorded five times and randomized to form an identification test.

This was done for each of E. N.'s fourteen pairs and for K. C.'s single pair, making fifteen tests of from thirty to fifty stimuli per test. The purpose of the carrier sentence was to provide a setting of the speaker's speech rhythm against which the subjects could judge the vowel durations as to length.

Since the vowels on the dynamic tones required special care in cutting, they were kept for testing until after the vowels on the static tones had been finished.

a. Vowels on static tones

Preserving the identity of static tones, when measured segments were cut from geminate vowels, presented no particular problem. A spectrogram was made of each altered utterance to be doubly sure of the duration of the segment removed and the placement of the cuts. This also insured that segments were cut out of steady-state portions of the vowel, thus preventing a discontinuity in formants with a resulting sudden shift in perceived quality. In addition, if a bad splice caused an audible click, the recording was discarded and a new copy used. In the one word with an initial vowel, /òot/, the cut was made after the beginning of the vowel, just as in those with initial consonants, so as to prevent the loss of the conditioned glottal stop.

Test 10

For this test K. C. recorded the sentence /dìcʻăn pʻûut kʻam wâa ˈhàat | hâj kʻun faŋ↓/ 'I (fem.) am saying the word /hàat/ (shoal) for you to listen to.' She also recorded the sentence with /hàt/ 'to practice'. Unaltered /hàt/ and /hàat/ together with eight experimental variants of /hàat/ are tabulated in terms of time removed and duration of the stimulus vowels. Vowel durations are given in milliseconds and as percentages of the duration of the geminate in /hàat/ to the nearest 1%.

TEST 10

VV = /hàat/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/hàat/:	None	360	100	100
'shoal'	50	310	86	100

TEST 10 — Continued

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
	75	285	79	100
	100	260	72	75
	125	235	65	51.7
	150	210	58	16.7
	175	185	51	0
	200	160	44	0
	225	135	38	0
/hàt/: ‘to practice’	None	130	36	0*

Number of subjects: 12 Total responses: 598

* I.e., all responses were /hàt/.

These ten stimuli were copied on tape five times, making fifty test items, which were scrambled to form a listening test. Twelve subjects were instructed to identify the key word in each occurrence of the carrier sentence as /hàt/ or /hàat/ and write it in Thai script. The results of the test are given in the table as percentages of the times the subjects reported hearing the geminate vowel.

Tests 11-24

For all the rest of the tests, including the ones on the dynamic tones, E. N. recorded the fourteen minimal pairs in the carrier sentence /pʻǒm pʻûut kʻam wâa ʻ______ | haj kʻun faŋ↓/ ‘I (male) am saying the word ______ for you to listen to.’ The two unaltered members of each pair as well as the shortened variants of the geminate vowel were copied on tape five times to form an identification test. On each occurrence of the carrier sentence, the subjects were asked to identify the key word as one of the pair and write it in Thai script.

The stimuli used in Tests 11-24 are described in tables like the one presented for Test 10. The tables also give the test results. A graphic display of these results is given in a composite figure at the end (Fig. 2.5).

TEST 11

VV = /bàat/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/bàat/: ‘Baht’	None	225	100	100
	50	175	78	100
	75	150	67	100
	100	125	56	81.7
	125	100	44	11.7
	150	75	33	0
	175	50	22	0
	200	25	11	0
/bàt/: ‘card’	None	100	44	0

Number of subjects: 12 Total responses: 540

TEST 12

VV = /cìip/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/cìip/: ‘to fold’	None	170	100	98.3
	50	120	71	65
	75	95	56	10
	100	70	41	3.3
	125	45	27	0
/cìp/: ‘to taste’	None	60	35	0

Number of subjects: 12 Total responses: 360

TEST 13

VV = /hàat/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/hàat/: 'shoal'	None	200	100	100
	50	150	75	100
	75	150	63	93.3
	100	100	50	18.3
	125	75	38	0
	150	50	25	0
/hàt/: 'to practice'	None	60	30	0

Number of subjects: 12

Total responses: 420

TEST 14

VV = /k‘ùut/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/k‘ùut/: 'to rub off'	None	210	100	100
	50	160	76	90
	75	135	64	78.3
	100	110	52	88.1
	125	85	41	0
	150	60	29	0
	175	35	17	0
/k‘ùt/: 'to dig'	None	60	29	0

Number of subjects: 12

Total responses: 479

TEST 15

VV = /baaŋ/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/baaŋ/: ‘some’	None	220	100	100
	50	170	77	58.3
	75	145	66	21.7
	100	120	55	8.3
	125	95	43	0
	150	70	32	0
	175	45	21	0
/baŋ/: ‘to hide’	None	100	46	6.7

Number of subjects: 12 Total responses: 480

TEST 16

VV = /òot/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/òot/: ‘to lament’	None	230	100	100
	50	180	78	100
	75	155	67	98.3
	100	130	57	38.3
	125	105	46	0
	150	80	35	0
	175	55	24	0
/òt/: ‘to abstain’	None	100	44	0

Number of subjects: 12 Total responses: 480

TEST 17

VV = /tʻaan/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/tʻaan/: ‘donation’	None	200	100	100
	50	150	75	100
	75	125	63	35
	100	100	50	1.7
	125	75	38	0
	150	50	25	0
/tʻan/: ‘on time’	None	70	35	0

Number of subjects: 12 Total responses: 420

TEST 18

VV = /hèet/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/hèet/: ‘reason’	None	210	100	100
	50	160	76	100
	75	135	64	78.3
	100	110	52	21.7
	125	85	41	0
	150	60	29	0
/hèt/: ‘mushroom’	None	90	43	0

Number of subjects: 12 Total responses: 420

TEST 19

VV = /pʻæǽ/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/pʻæǽ/: 'to be defeated'	None	330	100	100
	50	280	85	100
	75	255	77	100
	100	230	70	92.3
	125	205	62	60
	150	180	55	13.8
	175	155	47	3.1
	200	130	39	0
/pʻǽ/: 'goat'	None	150	46	0

Number of subjects: 13 Total responses: 585

TEST 20

VV = /pʻraá/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/pʻraá/: 'jungle knife'	None	300	100	100
	50	250	83	100
	75	225	75	100
	100	200	67	93.8
	125	175	58	36.9
	150	150	50	1.5
	175	125	42	0
	200	100	33	0
/pʻrá/: 'priest'	None	120	40	0

Number of subjects: 13 Total responses: 585

TEST 21

VV = /k'áan/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/k'áan/: 'to hinder'	None	200	100	100
	50	150	75	100
	75	125	63	91.7
	100	100	50	21.7
	125	75	38	0
	150	50	25	0
/k'án/: 'to press'	None	90	45	0

Number of subjects: 12 Total responses: 420

b. Vowels on dynamic tones

Because of the sharp upward and downward movement of the voice pitch in the dynamic tones, it was thought at first that it would be difficult or even impossible to cut successively longer segments from the vowels and still preserve the identity of the tones; that is, it seemed that so much of the contour would be cut out for the shortest variants that the resulting utterances would be unnatural or that another tone would replace the intended one. An available solution would have been to combine synthesis of tones of suitable lengths on the Intonator with resynthesis of the altered and unaltered vowels of real speech. This procedure, of course, would have involved a departure from the design of the experiments on the vowels on static tones and would have been a step further away from real speech.

By this point in the study, considerable analysis of the fundamental frequency patterns of the tones had been made for Chapter III. These measurements showed that even if as much as 75% of the duration of a geminate vowel were removed, about half of the fall of the falling tone or the rise of the rising tone would remain. Inasmuch as manifestations of these tones in connected speech often seem to be little more than enough of a fall or rise, respectively, to distinguish the occurring tone from all others in the system, it was felt that the cuts might be made without untoward consequences and that recourse to partly synthetic speech would then be unnecessary. The resulting variants ought to be acceptable to native speakers as utterances on the dynamic tones.[4] This turned out to be so. Upon being asked, all subjects readily responded that the tones were "normal".

The only change in the method was to make the first cut as near the beginning of the vowel as possible without impairing transitions, so as not to have to cut too far into the tonal contours. Using magnetic clicks for more precise cutting (Introd. C-1) might have interfered with the carrier sentence, so audible stop bursts on the magnetic tape and corresponding points on spectrograms were used instead. Wide-band spectrograms were made of all variants to check the accuracy of the cutting and splicing.

TEST 22

VV = /kʻâam/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/kʻâam/: 'to cross'	None	200	100	100
	50	150	75	98
	75	125	63	54
	100	100	50	0
	125	75	38	0
/kʻâm/: 'dusk'	None	80	40	0

Number of subjects: 10

Total responses: 300

TEST 23

VV = /sǐin/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/sǐin/: ‘moral code’	None	160	100	100
	25*	135	84	100
	50	110	69	60
	75	85	53	40
	100	60	38	1.8
/sǐn/: ‘riches’	None	60	38	0

Number of subjects: 11 Total responses: 330

* The geminate vowel in this word was relatively short (Table 2.4), and making a first cut of 50 msec would have meant a jump from 100% to less than 70%, out of keeping with the percentage changes in all the preceding tests; consequently, the first portion removed was only 25 msec long. This was also done in Test 24.

TEST 24

VV = /kʻǎaj/

	Time Removed (msec)	Stimulus Duration (msec)	Duration as Percent VV	Percent VV Responses
/kʻǎaj/: ʻto sellʼ	None	160	100	100
	25	135	84	100
	50	110	69	100
	75	85	53	69.1
	100	60	38	7.3*
/kʻǎj/: ʻto windʼ	None	80	50	0

Number of subjects: 11 Total responses: 330

* It is interesting to note that this variant was shorter than the original V. The three subjects responsible for these five judgments may have been influenced by the duration of the final /j/. For an explanation of the treatment of the semivowels in these syllables, see Sec. A-3.

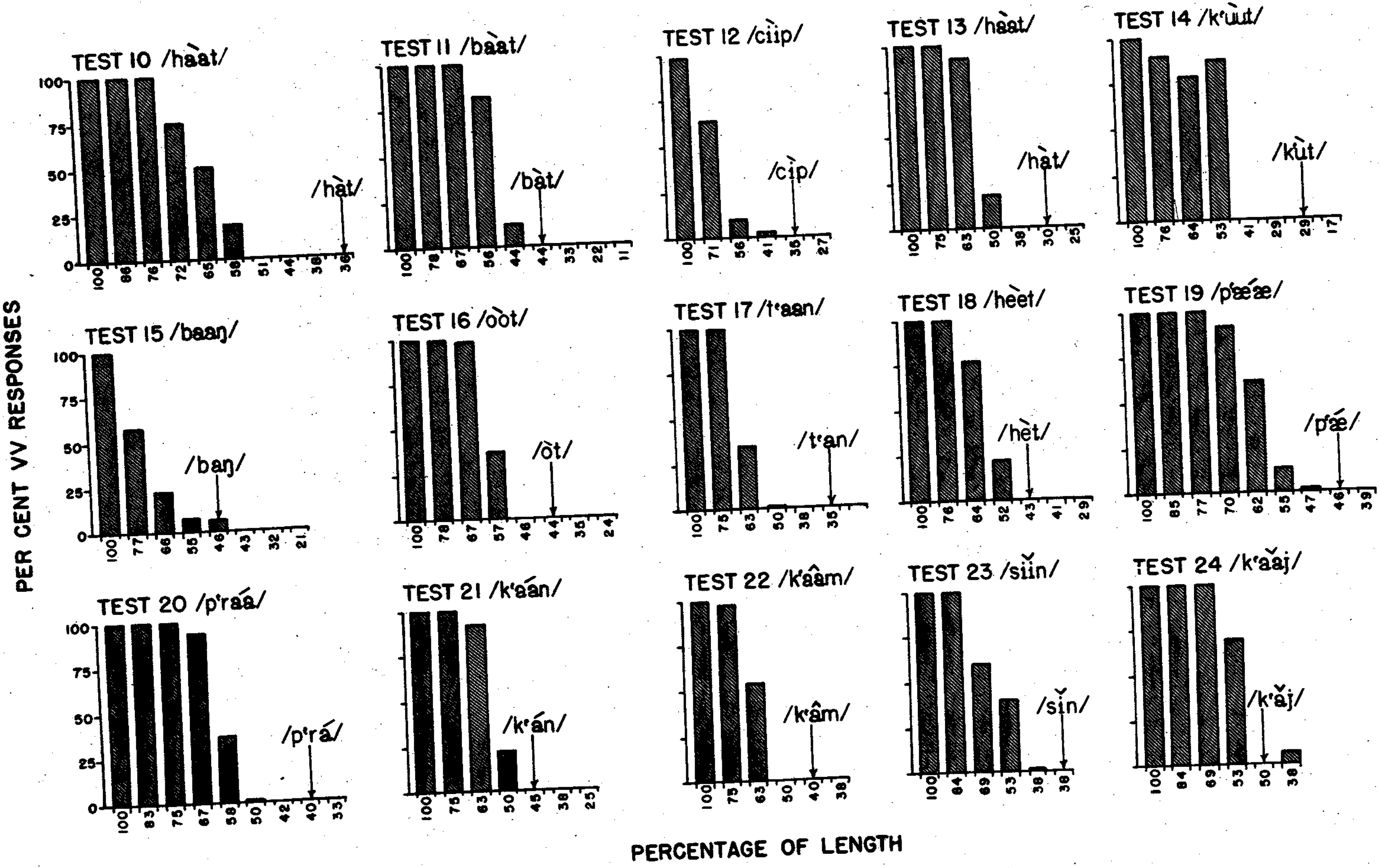


FIG. 2.5 PERCENTAGE VV RESPONSES IN TESTS 10-24

C. Discussion of Results

1. Vowel Duration and Tones

As a link with the study of tonal features, it should be of some value to see whether sub-phonemic differences of vowel duration are concomitant with the five tones. The average durations of single and geminate vowels occurring on each tone were calculated from the data of Table 2.1. These averages are given for the two speakers, W. N. and E. N., in Table 2.5.

TABLE 2.5

Tones Correlated with Average Vowel Durations in Msec

Tone	V		VV	
	W. N.	E. N.	W. N.	E. N.
/0/			492	404
/\`/	163	158	524	413
/´/	97	165	436	352
/^/			454	343
/ˇ/			478	371

Since citation forms of single vowels are limited to the low and high tones (Introd. B-3d), it is impossible to draw any conclusions as to correlations between allophonic length variations and tones from the restricted data of Table 2.5; for W. N., /V̀/ is longer than /V́/, while the reverse is true for E. N. The geminate vowels, however, show quite a good correlation for the two speakers, as can be seen in Table 2.6, even though the deviations from the mean durations are not large.

TABLE 2.6

Tones in Rank Order of Average Durations of Geminate Vowels

Order	W. N.	E. N.
1	/ˋ/	/ˋ/
2	/0/	/0/
3	/ˇ/	/ˇ/
4	/ˆ/	/ˊ/
5	/ˊ/	/ˆ/

The falling and high tones are set off against the other three, which have the same order for both speakers. That the situation is much the same for E. N.'s vocalic clusters is shown by Table 2.7, which gives the average durations of the clusters on each tone in Table 2.2 and the resulting rank ordering of the tones.

TABLE 2.7

Vocalic Clusters: Average Durations and Tones

Tone	Msec	Rank Order
/0/	487	1
/ˋ/	487	1
/ˊ/	373	4
/ˆ/	420	3
/ˇ/	460	2

Generalizing from the last two tables, it can at least be said that the mid and low tones tend to be longer, and the high and falling tones shorter, than the rising tone. With these findings as a point of departure, such correlations should be investigated in running discourse. Although the minimal pairs (Sec. A-3) were recorded under the same syntactic and prosodic conditions, for the present purpose they do not offer enough examples of vowels on the same tone in the same or analogous consonantal environments.

2. Primacy of the Quantitative Feature

The primacy of relative duration as the feature that distinguishes "short" vowels from "long" vowels has been thrown into question by Marguerite Durand, who advances the hypothesis that the perception of duration depends on the stimulation caused by the various characteristics of long or short vowels. "La persistance de la réponse physiologique explique l'impression de durée donnée par les longues, quelle que soit leur mesure réelle."[5] The phonetic characteristics distinguishing long and short vowels, she says, are differences of fortisness, intensity, muscular tension, breath flow, pitch movement, linkage with the following consonant, quality, and duration (pp. 162, 178). Duration as such, in her view, has little if anything to do with perception of the length distinction. In arriving at her conclusions, Durand utilized not only her own investigation of various languages but also the work of others, including a study on vowel duration in Thai.[6]

It is believed that the experiments reported in the present study refute Durand's thesis, at least insofar as Standard Thai is concerned. It seems clear that the cue of relative duration alone serves to distinguish single vowels from geminates in Thai, although other features, including some or even all of the ones described by Durand, probably have redundant value in this contrast. These features have not been investigated experimentally, but, except for a few that may have been affected by the artificial nature of the changes in duration, most of her causative factors were left unaltered and free to counteract temporal differences. Indeed, two features heavily emphasized by Durand, variations in the final portions of phonemic tones and close or loose contact with the following consonant, were left completely untouched.

As clearly shown by the bar graphs in Figure 2.5, subjects can be led to perceive originally geminate vowels embedded in a carrier sentence as single vowels by shortening the duration past a certain point. There is a brief transitional zone of ambiguity on either side of which vowels are predominantly long or short, that is, geminate or single. To find zones of perceptual uncertainty for the vowels in the minimal pairs, the results of Tests 10-24 were replotted as points on graph paper with the same coordinates as in the bar graphs. The "crossover point" (i.e., the center of the zone of ambiguity) in each test was taken to be the intersection of a smooth curve drawn through the points and the 50% response line. The crossover points are given in Table 2.8 as percentages of the durations of the original geminates and as durations in milliseconds. The durations of the single vowels are given for comparison.

TABLE 2.8

Crossover Points between Single and Geminate Vowels

Test	Word	Crossover Point Percent	Msec	VV Duration in Msec	V Duration in Msec
10	hàat	65	234	360	130
11	bàat	51	115	225	100
12	cìip	66	112	170	60
13	hàat	55	110	200	60
14	kʻùut	47	99	210	60
15	baaŋ	74	163	220	100
16	ɔ̀ot	57	131	230	100
17	tʻaan	64	128	200	70
18	hèet	59	124	210	90
19	pʻǽæ	61	201	330	150
20	pʻráa	59	177	300	120
21	kʻáan	53	106	200	90
22	kʻâam	61	122	200	80
23	sǐin	61	98	160	60
24	kʻǎaj	48	77	160	80
Average:[7]		58	126	215	87

Looking at the above data, we see that there is a large difference in duration between V and VV and that subjects are able to give crossover points at an intermediate point. These findings strongly support the conclusion that the feature of relative length serves quite adequately to distinguish single vowels from geminate ones.

Notes

1. The experimenter did not request /ˈ/ and /|/, but the informant supplied them in all recordings of the sentence.

2. The vocalic clusters /ia ɨa ua/, which together with the geminates comprise the class of double vowels, are obviously identifiable by their timbre.

3. Introduction, B-2a. Allophones and Figs. 1.10-11.

4. Experiments on the perception of tonal features are described in Chap. III. This study has not been extended to stripping the tones down to the minimum cues needed for intelligibility.

5. M. Durand, *Voyelles longues et voyelles brèves* (Paris, 1946), p. 185.

6. Kasem Sibunruang, *Essai sur l'étude de la durée vocalique en siamois*. Mém. de l'Institut de Phonétique. Paris: 1940. Extensive efforts to obtain a copy of this unpublished ms. have failed. Durand summarizes it along with other important works on vowel length; see her bibl.

7. These averages represent Tests 11-24, which were based on the productions of one speaker. See the entry for Test 10 for the one other speaker used. There is good correlation between these values and the measurements of the citation forms (Table 2.1).

III. TONES

The five phonemic tones of Standard Thai are analyzed into two dynamic tones and three static tones (Introd. B-2c). The first group is characterized by sharp pitch contours as opposed to the relatively smooth pitch movements of the second. Indeed, the static tones sound level or nearly so in some environments, while the dynamic tones never do. The physical parameter usually associated with change of pitch is variation in fundamental frequency. This chapter deals with measurements of fundamental frequency, their correlation with the pitch patterns of the phonemic tones, and the perceptual primacy of the fundamental frequency in signalling the tonal distinctions. Brief consideration is also given to the emphatic tone.

A. Measurements

1. Tones on Monosyllabic Morphemes

In keeping with the practice of the preceding chapters, citation forms were examined to determine idealized tonal shapes. Eighty-eight monosyllabic morphemes, thirty-nine on single vowels and forty-nine on double vowels, were recorded on tape by E. N. The list of morphemes comprised sets of two, three, four or five syllables minimally distinguished by tone. The sets are given in Table 3.1.

TABLE 3.1

Morphemes Used for Tonal Measurements

Single Vowels

/0/	/\`/	/´/	/ˆ/	/ˇ/
kʻaj	kʻàj	kʻáj	kʻâj	kʻǎj
(ʻdried sweat')	(ʻegg')	(ʻto scoop out')	(ʻfever')	(ʻto unlock')

TABLE 3.1 — Continued

/0/	/`/	/´/	/^/	/ˇ/
kʻun ('you')	kʻùn ('muddied')	kʻún ('familiar')	----	kʻŭn ('to feed')
lom ('wind')	lòm ('pit')	lóm ('to fall')	lôm ('shipwreck')	----
kʻam ('word')	----	kʻám ('to support')	kʻâm ('dusk')	kʻăm ('funny')
----	kʻàw ('knee')	kʻáw ('token')	kʻâw ('to enter')	kʻăw ('he')
pʻɨŋ ('ought')	pʻɨ̀ŋ ('to spread')	----	pʻɨ̂ŋ ('to depend on')	pʻɨ̆ŋ ('sound of breaking rope')
----	kʻæ̀ŋ ('to race')	----	----[1]	kʻæ̆ŋ ('hard')
----	pə̀n ('awkward')	----	pə̂n ('he')	----
----	wàt ('a cold')	wát ('monastery')	----	----
----	nàk ('heavy')	nák ('expert')	----	----
----	sɔ̀ŋ ('to project a light')	----	sɔ̂ŋ ('slum')	----
----	pʻìt ('different')	pʻít ('poison')	----	----
----	pʻèt ('spicy')	pʻét ('diamond')	----	----
		Double Vowels		
naa ('field')	naà ('custard apple')	naá ('mother's younger sibling')	naâ ('face')	naǎ ('thick')
loo ('unstable')[2]	loò ('least')	loó ('to swing')	loô ('shield')	loǒ ('dozen')

TABLE 3.1 — Continued

/0/	/ˋ/	/ˊ/	/ˆ/	/ˇ/
tʻii ('instance')	tʻìi ('right after')	----	tʻîi ('place')	tʻǐi ('woman')
tʻuu ('Rastrelliger chrysozonus')[3]	----	tʻúu ('to balance')	tʻûu ('blunt')	tʻǔu ('to rub')
luaŋ ('to deceive')	----	lúaŋ ('to insert')	lûaŋ ('to pass time')	lǔaŋ ('Luang (a title)')
pʻææ ('raft')	pʻæ̀æ ('to spread')	pʻǽæ ('to be defeated')	----	----
----	sɨ̀a ('mat')	----	sɨ̂a ('upper garment')	sɨ̌a ('tiger')
----[4]	kʻàaw ('news')	----	kʻâaw ('rice')	kʻǎaw ('white')
sɔɔŋ ('envelope')	----	sɔ́ɔŋ ('to praise in chorus')	----	sɔ̌ɔŋ ('two')
mɔɔ ('gray')	----	----	mɔ̂ɔ ('pot')	mɔ̌ɔ ('healer')
əəj ('(final word in verse)')	ə̀əj ('to mention')	----	----	ə̌əj ('(term of endearment)')
hee ('hey')	hèe ('lullaby')	----	----	hěe ('to turn aside')
cʻaa ('tea')	----	cʻáa ('slow')	----	----
----	----	sɨ́ɨ ('to buy')	sɨ̂ɨ ('honest')	----
tʻiaw ('exactly')	----	----	tʻîaw ('to roam')	----

Narrow-band spectrograms on both the 1200 and the 200 cycle scales were made of all the syllables. Examples of the former are shown in Figure 3.1, and of the latter, in Figure 3.2. The same syllables appear in the two figures. Because of their striking display of the course of the lower harmonics, the 200 c/in. spectrograms were used, with proper calibration, to trace the fundamental frequency patterns of the tones. When, on occasion, the width of the harmonic streaks in this display caused uncertainty, the strong higher harmonics on 1200 c/in. spectrograms were measured as an additional check. The latter were also useful for general orientation within the syllable.

Lines perpendicular to the time axis were drawn across the first two harmonics at intervals of twenty-five milliseconds for rapidly moving stretches and fifty milliseconds or more for slowly moving stretches and level portions. The space between the centers of the harmonic streaks was measured along each perpendicular line. This distance was laid off against the relevant portion of the calibration curve[5] to find the fundamental frequency of the voice at that instant. Wherever the fundamental was obscure or absent, higher harmonics were measured to obtain the fundamental frequency.

Because the presence of a final double vowel, semi vowel, or nasal permits the occurrence of any of the tones (Introd. B-3d), the domain of the tone was taken to be the syllable rather than just the vowel, and measurements were made from the beginning of observable periodicity to the end. In this way tonal contours were tracked to their completion.

In some of the spectrograms of syllables on the high tone, e.g. /kʻún/ in Figure 3.2, a break in the harmonics near the end corresponds to audible glottal constriction. The breathy voice following this appears to be produced with much the same glottal tension; consequently, intermittent bursts of tonality that occur may appear to be continuations of the harmonics found before the glottalization, but they have not been included in the measurements.[6]

The values obtained for each syllable were plotted on a graph of frequency versus time, and the points were connected to form a tonal curve. As an illustration of this, the curves plotted for the rising tone on syllables with single vowels are shown in Figure 3.3. The upper limit for all of E. N.'s tones, including the emphatic tone, was 165 cps, and the lower limit was 83 cps.

All tonal measurements of syllables with single vowels are entered in Table 3.2 and double vowels in Table 3.3.[7] Each table is subdivided into tones. The values are taken from the raw curves every twenty-five milliseconds. A number in parentheses represents

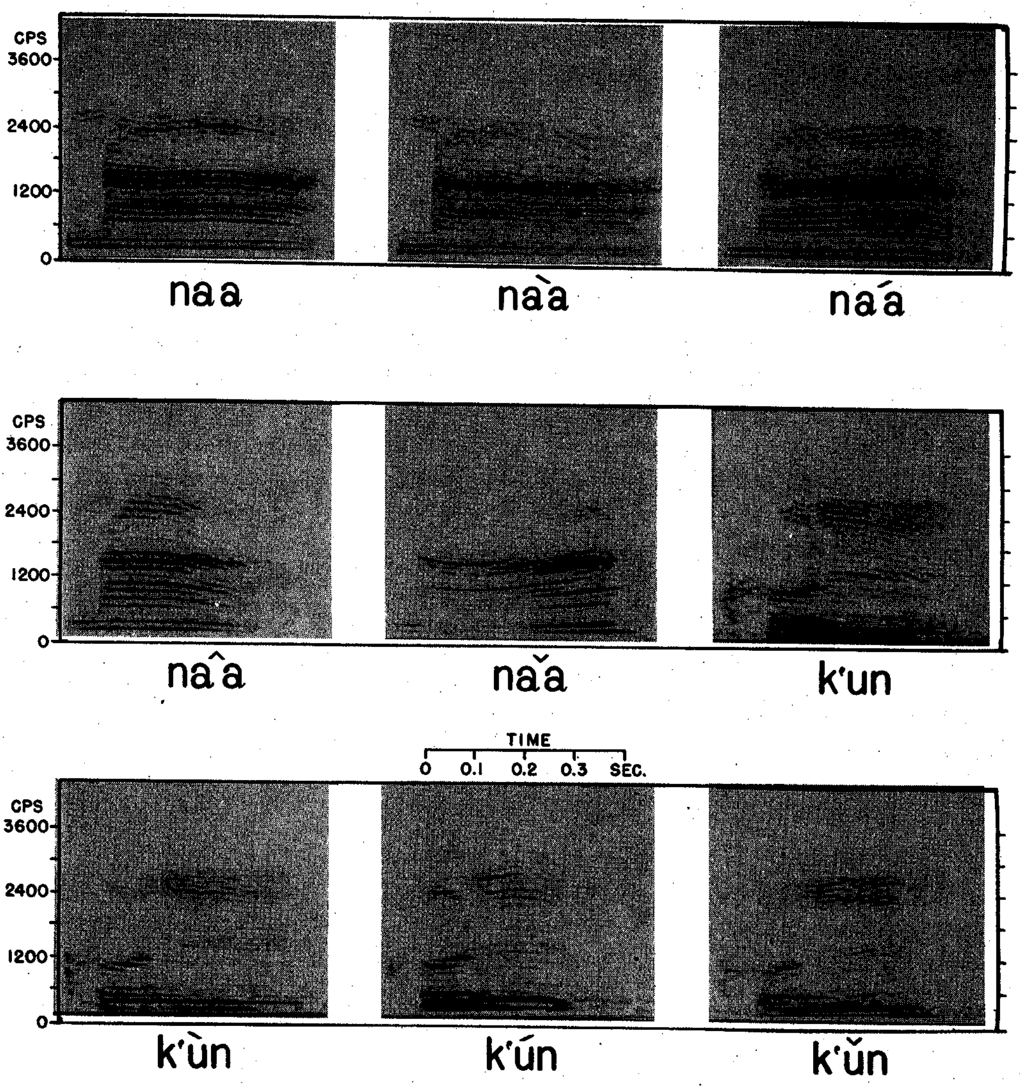


FIG. 3.1 NARROW-BAND SPECTROGRAMS OF 2 SETS OF TONALLY DIFFERENTIATED MORPHEMES.

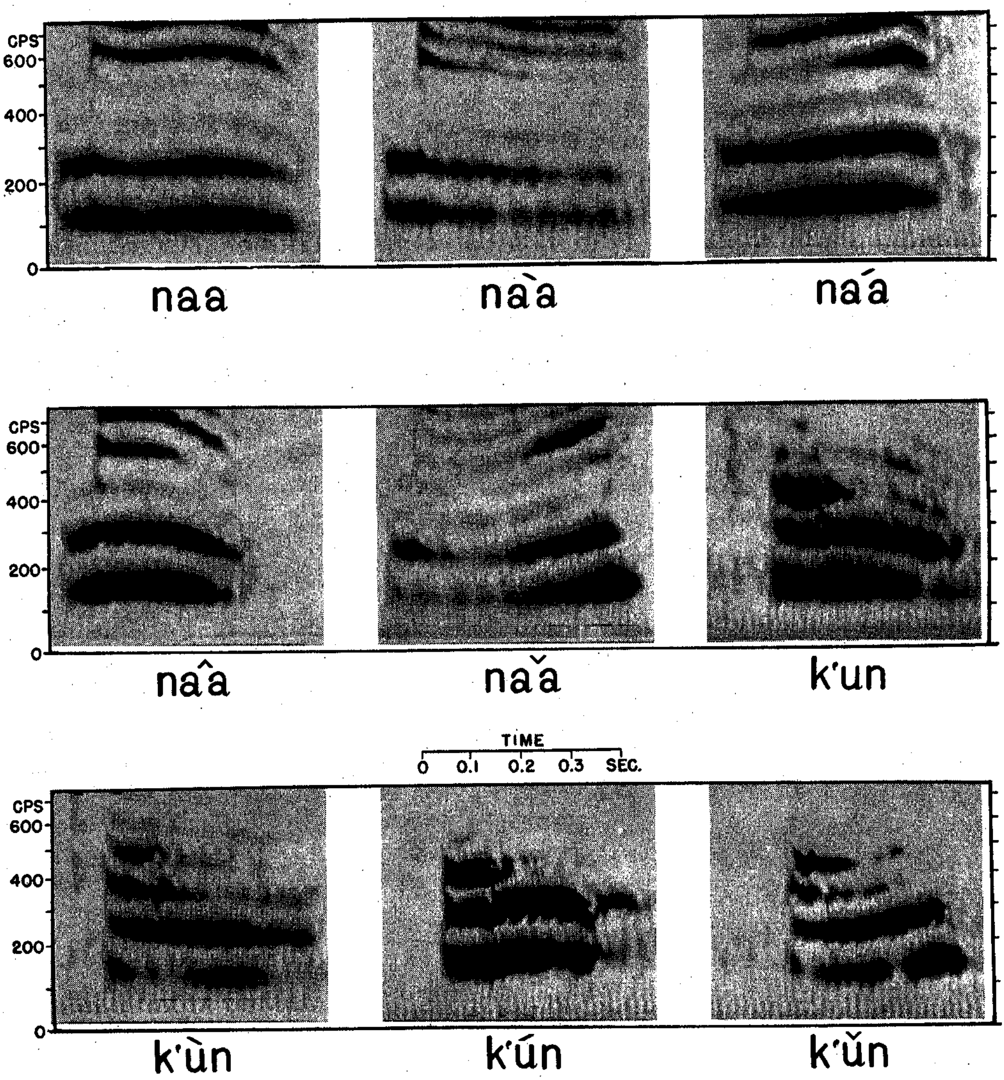


FIG. 3.2 LOWER HARMONICS OF 2 SETS OF TONALLY DIFFERENTIATED MORPHEMES.

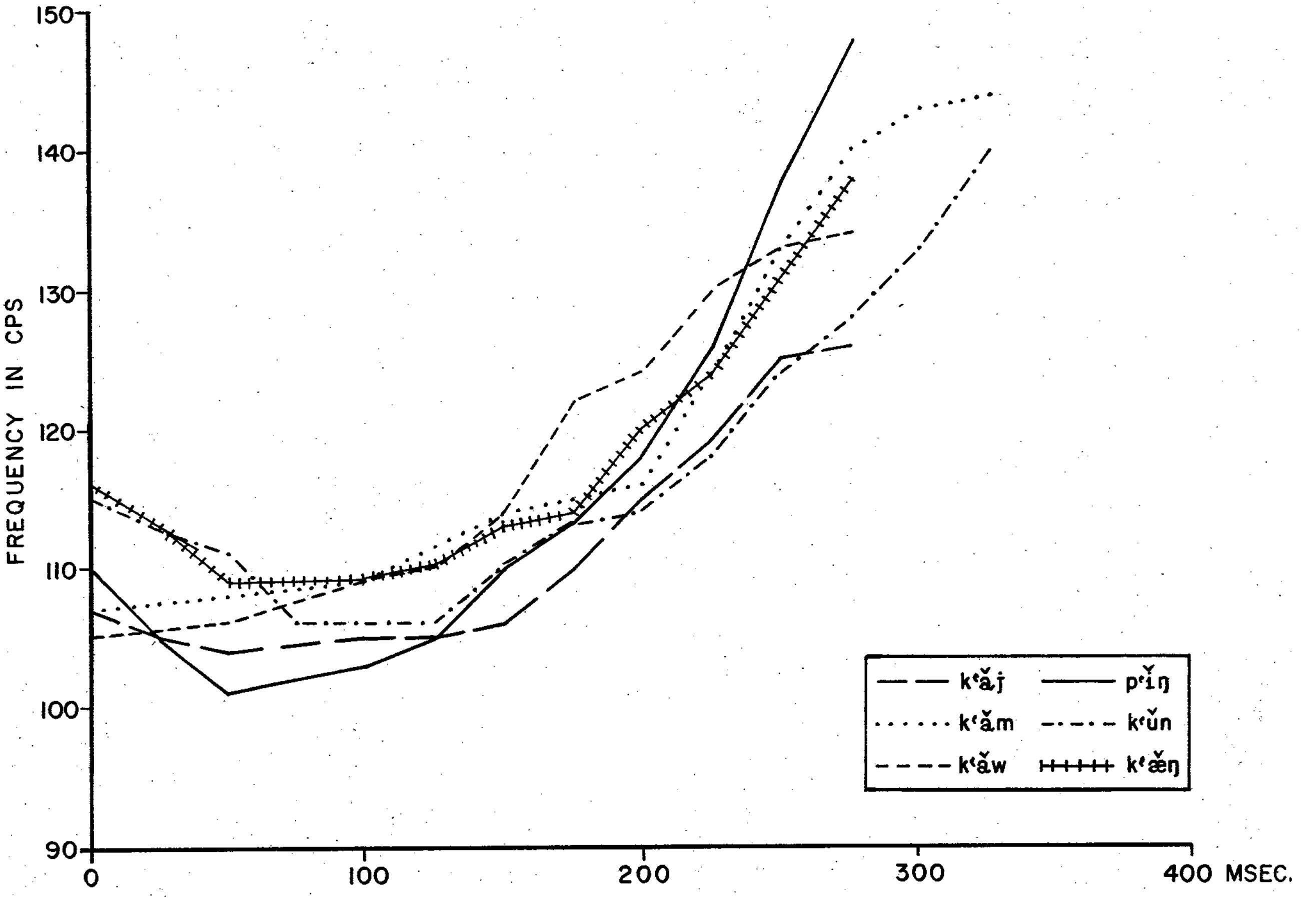


FIG. 3.3 RISING TONE ON SINGLE VOWELS.

a linear interpolation from the curve rather than an actual measurement at that point, as can be seen by comparing Figure 3.3 with Table 3.2E. If the last measurement was not made at a multiple of twenty-five milliseconds, it is entered only in the column of Terminal Values. Otherwise, it is put both there and in the proper column within the table. All the raw curves can be reconstructed simply by plotting the values in the tables.

In order to arrive at average tonal curves, it was first necessary to normalize the raw curves as to duration. This was done by shrinking or stretching the curves proportionally to a convenient length. Syllables with double vowels were stretched to 500 msec, a duration slightly longer than the largest measured value. Since syllables with single vowels varied considerably in length, at least on the low and high tones, a compromise suitable for graphic display and reasonably representative of the measured durations was effected by normalizing them to half the duration of the double vowels.

An average curve was then drawn through the family of normalized curves of each plot. This was done by eye as an easy and reasonable way of achieving weighting. It is believed that repetitions of the procedure would yield curves of essentially the same shape with only minor variations in absolute value. As an illustration of the above procedures, the raw curves of measured values for the rising tone on single vowels (Fig. 3.3 and Table 3.2E) have been normalized and averaged in Figure 3.4.

The average curves for single vowels are presented in Figure 3.5 and for double vowels in Figure 3.6. Before a final stop or zero, the high tone on a single vowel does not drop at the end; this major variant is indicated by a broken line in Figure 3.5. A small sample[8] of the speech of a female informant, K. S., reveals the same general tonal configurations, although her pitch range of 175 cps to 315 cps is completely above that of E. N.

2. The Emphatic Tone

Although this study is concerned with vocalic and tonal phonemes, one emphatic feature (Introd. B-2e), the emphatic tone, was examined briefly. Four verbal reduplications were uttered by E. N. with the emphatic tone on the first morph of each and recorded on tape:

di̋idii↓ 'How good it is!'
rɔ̋ɔnrɔɔ́n↓ 'How hot it is!'
wa̋anwǎan↓ 'How sweet it is!'
ja̋akjâak↓ 'How difficult it is!'

TABLE 3.2

TONAL CONTOURS OF SINGLE VOWELS (CYCLES PER SECOND)

Msec:	0	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	Terminal	Value
																				Pitch	Time
A. Mid Tone																					
kʻaj	118	115	115	(115)	115	(115)	115	(115)	114	(114)	(114)	(113)	111	110						110	325
kʻun	124	(122)	120	(123)	126	(127)	127	120	119	118	113	109								109	275
lom	120	(120)	120	(120)	120	(120)	120	(121)	121	(121)	121	120	115	106	104	103				103	375
kʻam	137	(137)	137	(137)	136	(136)	136	(136)	136	(130)	125	(125)	125							125	300
B. Low Tone																					
kʻàj	111	(109)	107	(106)	105	(105)	104	(104)	104	(104)	104	(105)	105	(105)	105	105				105	375
kʻùn	120	115	(114)	(113)	111	(111)	110	(109)	(107)	(106)	105	(105)	104	(103)	102	(100)	97			97	400
lòm	122	(118)	115	(114)	113	(112)	110	(109)	108	(108)	108	(108)	107	(107)	106	(106)	106	(106)		106	440
kʻàw	111	110	110	106	105	(105)	105	(105)	(105)	(105)	105	(105)	(105)	(105)	105					105	350
pʻɨ̀ŋ	120	113	110	106	104	98	92	(92)	92	(92)	(92)	(92)	92	(92)	92	(92)	92			92	400
kʻæ̀ŋ	124	115	113	(112)	111	(110)	108	(108)	108	(107)	106	105	104	(105)	105					105	370
pə̀n	132	120	114	113	111	(109)	107	(106)	105	(105)	105	(105)	105	(105)	105	(106)	106	(106)	106	106	450
wàt	118	115	114	(114)	113															111	115
nàk	121	115	108	105	103	103	101													101	150
sɔ̀ŋ	123	(119)	114	(114)	113	(112)	110	(110)	110	(110)	110	(110)	110	(110)	110	110				110	375
pʻìt	131	127	125	118	112															112	100
pʻèt	120	118	118	116																116	75
C. High Tone																					
kʻáj	117	(118)	118	(120)	121	124	127	127	130	(130)	130	123								122	295
kʻún	121	(121)	121	124	127	133	135	139	139	(136)	132	130								130	275
lóm	83	94	108	120	127	(129)	130	(133)	136	(138)	140	(135)	129	118						118	325
kʻám	141	(142)	142	145	156	159	159	156	154	149	141									138	270

TABLE 3.2 — Continued

Msec:	0	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	Terminal Pitch	Value Time
kʻáw	120	(120)	120	120	123	129	142	141	146	143	137	132								132	275
wát	120	127	127	(129)	131	(131)	131													131	170
nák	116	118	126	(129)	132	130	129	129												129	175
pʻít	139	(145)	150	(153)	155															155	115
pʻét	141	141	143	143	143															143	110
D. Falling Tone																					
kʻâj	137	142	133	132	127	125	118	114	104	102	96									96	250
lôm	130	132	138	138	134	128	122	115	110	102	93									93	250
kʻâm	140	153	153	157	151	149	144	130	116	110	107	106								106	275
kʻâw	147	148	146	148	141	126	119	106	100	99										99	225
pʻɨ̂ŋ	150	(150)	150	(145)	140	135	129	109	101	93										93	225
pə̂n	142	(140)	137	133	130	123	114	105	97	95	95									95	250
sɔ̂ŋ	130	124	123	118	118	118	115	105	100	96										96	225
E. Rising Tone																					
kʻǎj	107	105	104	(105)	105	105	106	110	115	119	125	126								126	275
kʻǔn	115	(113)	111	106	106	106	110	113	114	118	124	128	133	140						140	325
kʻǎm	107	(108)	108	(109)	109	(112)	114	115	116	124	133	140	143	144						144	325
kʻǎw	105	(106)	106	(108)	109	110	114	122	124	130	133	134								134	275
pʻɨ̌ŋ	110	105	101	(102)	103	105	110	113	118	126	138	148								148	275
kʻæ̌ŋ	116	113	109	(109)	109	110	113	114	120	124	131	138								138	275

TABLE 3.3

TONAL CONTOURS OF DOUBLE VOWELS (CYCLES PER SECOND)

Msec:	0	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	Terminal Pitch	Value Time
										A. Mid Tone											
naa	130	(132)	134	135	129	(129)	129	(130)	130	(131)	132	(132)	131	(130)	129	125	122	117		117	425
loo	122	(122)	122	(123)	124	(122)	120	(120)	120	(121)	122	(123)	124	(123)	122	(121)	120	116	116	116	450
tʻii	131	131	133	136	133	(133)	133	(135)	136	(135)	133	129	126	124	120					120	350
tʻuu	128	(130)	131	(127)	123	(124)	124	(124)	(124)	(124)	124	(122)	120	(121)	121	(120)	119	(119)	119	119	450
luaŋ	136	140	141	(141)	(141)	(141)	141	(141)	140	(139)	137	(137)	137	(137)	136	128	126	124		124	425
pʻææ	144	125	121	(121)	120	(121)	122	(122)	122	(124)	126	121	120	117						117	325
sɔɔŋ	123	120	124	(124)	123	(124)	124	(125)	125	(125)	125	(125)	125	(125)	125	125	121			121	400
mɔɔ	130	(130)	130	(130)	130	(130)	130	(130)	130	(130)	130	(130)	130	(130)	130	(129)	128	126		126	425
əəj	133	(133)	132	(130)	127	(126)	124	(125)	125	(125)	125	(125)	125	(124)	122	119	114	114		114	425
hee	117	(117)	117	118	119	(119)	119	(119)	119	(119)	108	(109)	109	109						109	325
cʻaa	137	129	119	(119)	119	(119)	119	(119)	119	(119)	119	(118)	116	(114)	112	102	95			95	400
tʻiaw	133	(134)	134	133	130	125	126	(126)	126	(126)	126	(126)	126	125	123	118	115	113		113	425
										B. Low Tone											
nàa	119	118	113	(112)	111	(111)	111	(109)	107	(106)	105	(104)	102	(102)	102	(101)	100	(100)	100	100	450
lòo	117	(116)	115	(114)	113	(113)	113	(112)	111	(110)	109	(109)	109	(108)	106	(106)	105	(105)	105	105	470
tʻìi	135	135	132	124	123	(121)	118	(116)	114	(114)	114	(114)	113	(112)	111					111	350
pʻæ̀æ	115	113	109	107	106	(105)	103	(103)	103	(103)	102	(102)	101	(101)	101					101	350
sìa	128	120	118	113	109	(109)	109	(109)	109	(108)	106	(106)	106	(106)	106	(106)	106			106	400
kʻàaw	117	114	115	115	112	(112)	112	(112)	112	(109)	105	(105)	105	(105)	105	(105)	105			105	400
ə̀əj	125	119	113	105	103	101	101	(101)	101	(101)	101	(101)	101	(101)	101	(101)	101			101	400
hèe	115	(114)	112	(111)	110	(109)	107	(107)	106	(106)	105	(105)	104	(103)	101	(101)	101			101	400

TABLE 3.3 — Continued

Msec:	0	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	Terminal Pitch	Value Time
C. High Tone																					
naá	102	115	122	126	132	(131)	129	(134)	139	142	145	(143)	140	(141)	141	141	138	139		133	435
loó	118	(119)	120	(122)	124	(125)	126	(126)	125	126	130	130	128	126						121	340
tʻuú	137	135	133	(133)	133	135	137	140	142	148	152	152	142	132						132	325
luáŋ	133	135	138	(141)	143	(144)	145	(148)	150	(152)	153	(154)	155	154	152					150	360
pʻæǽ	120	122	117	(121)	125	(126)	126	(128)	129	125	129	130	130							130	300
sɔɔ́ŋ	130	130	130	(130)	129	(130)	131	(135)	138	142	145	(145)	144	141	137	135				135	375
cʻaá	127	125	128	(128)	128	(128)	127	(128)	129	(132)	135	(136)	137	(138)	138	138	133			133	400
sɨɨ́	145	142	142	139	139	(142)	145	(144)	143	(144)	145	(147)	148	136	135					135	350
D. Falling Tone																					
nâa	110	118	130	131	(131)	131	(131)	131	(128)	124	(121)	117	(115)	112	(107)	101				98	385
lôo	130	133	138	141	141	140	137	132	127	120	113	108	101							101	300
tʻîi	160	(160)	160	159	154	152	148	138	128	118	112	103								98	285
tʻûu	145	(145)	145	145	137	131	126	120	110	105	100	98								98	275
luâŋ	133	138	143	148	150	151	148	146	143	138	132	123	114	103						103	325
sɨ̂a	150	(149)	148	148	148	145	139	135	130	115	110	105	100	95						95	325
kʻâaw	159	(156)	153	(153)	153	(151)	149	149	135	130	122	112	110	101	98	90				90	375
mɔ̂ɔ	146	154	158	(157)	156	(156)	156	(152)	147	140	135	132	118	100	94					94	350
sɨ̂ɨ	155	(155)	155	(153)	150	148	143	136	135	128	122	120	113	109						109	325
tʻîaw	155	164	163	154	148	148	147	137	119	105	100	95	93							93	300

TABLE 3.3 — Continued

Msec:	0	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	Terminal Pitch	Value Time
									E. Rising Tone												
nǎa	118	110	108	105	104	102	100	(102)	104	(106)	107	(109)	111	113	119	123	125	133	140	140	450
lǒo	111	112	112	(111)	110	(110)	110	(111)	112	111	111	116	122	127	131	141	145			145	400
tʻǐi	115	115	112	(109)	106	(108)	110	(115)	119	(125)	130	135	135	142	152	153				153	375
tʻǔu	124	114	112	108	108	110	112	114	117	124	128	133	142	152	160					160	350
lǔaŋ	120	(121)	122	(121)	120	(118)	115	114	114	116	119	124	132	140	147	148				148	375
sǐa	103	108	107	(108)	108	(109)	110	(109)	108	115	119	125	131	141	142	142				142	375
kʻǎaw	107	(108)	108	(109)	110	114	115	119	125	128	133	144	148	158						158	325
sɔ̌ɔŋ	93	100	102	(102)	102	(102)	102	(111)	115	113	123	122	127	130	140	146	153			153	400
mɔ̌ɔ	113	(112)	111	(109)	106	(106)	105	(105)	105	(106)	106	110	115	119	123	130	133	143	150	150	450
ə̌əj	124	119	113	110	107	106	105	(106)	106	110	113	118	122	125	132	137	143	153		153	425
hěe	128	120	111	(111)	(111)	111	112	116	117	117	120	121	126	134	137	141				141	375

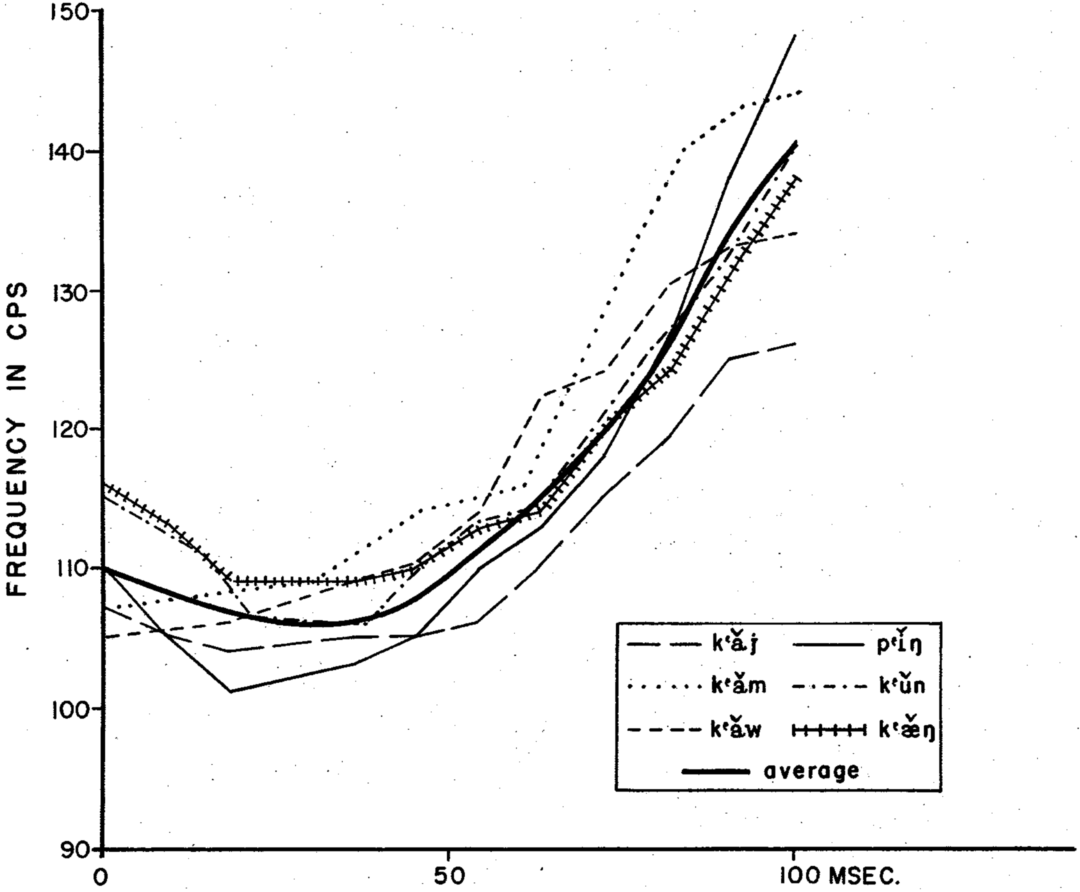


FIG. 3.4 RISING TONE ON SINGLE VOWELS: NORMALIZED CURVES AND AVERAGE CURVE.

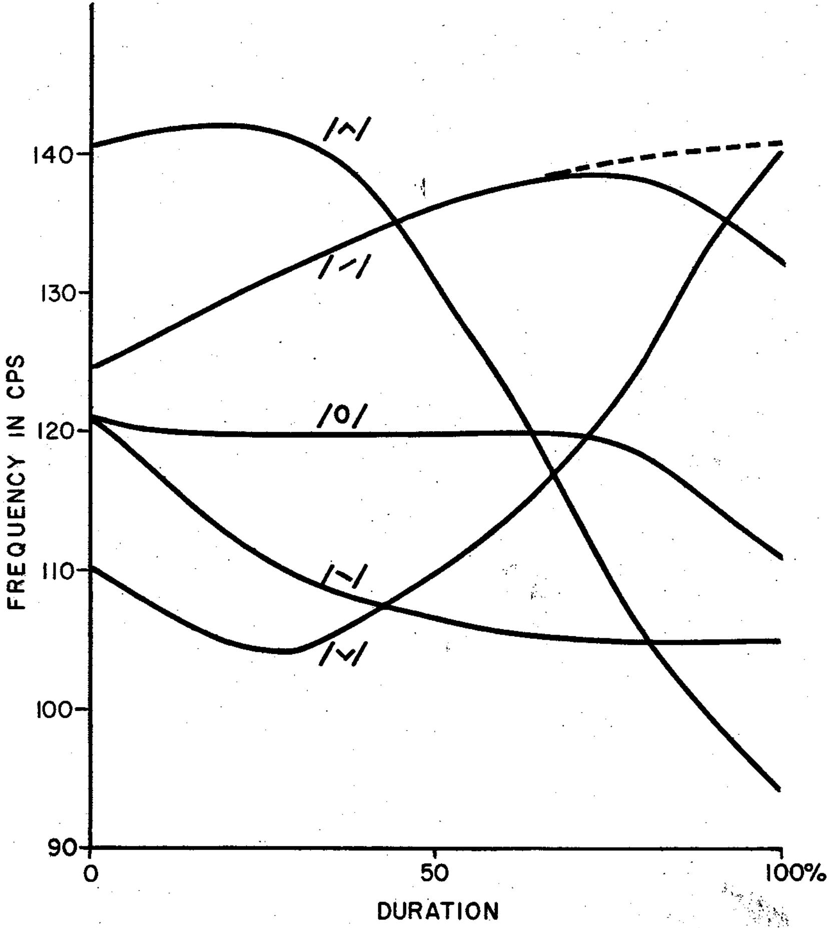


FIG. 3.5 TONES ON SINGLE VOWELS

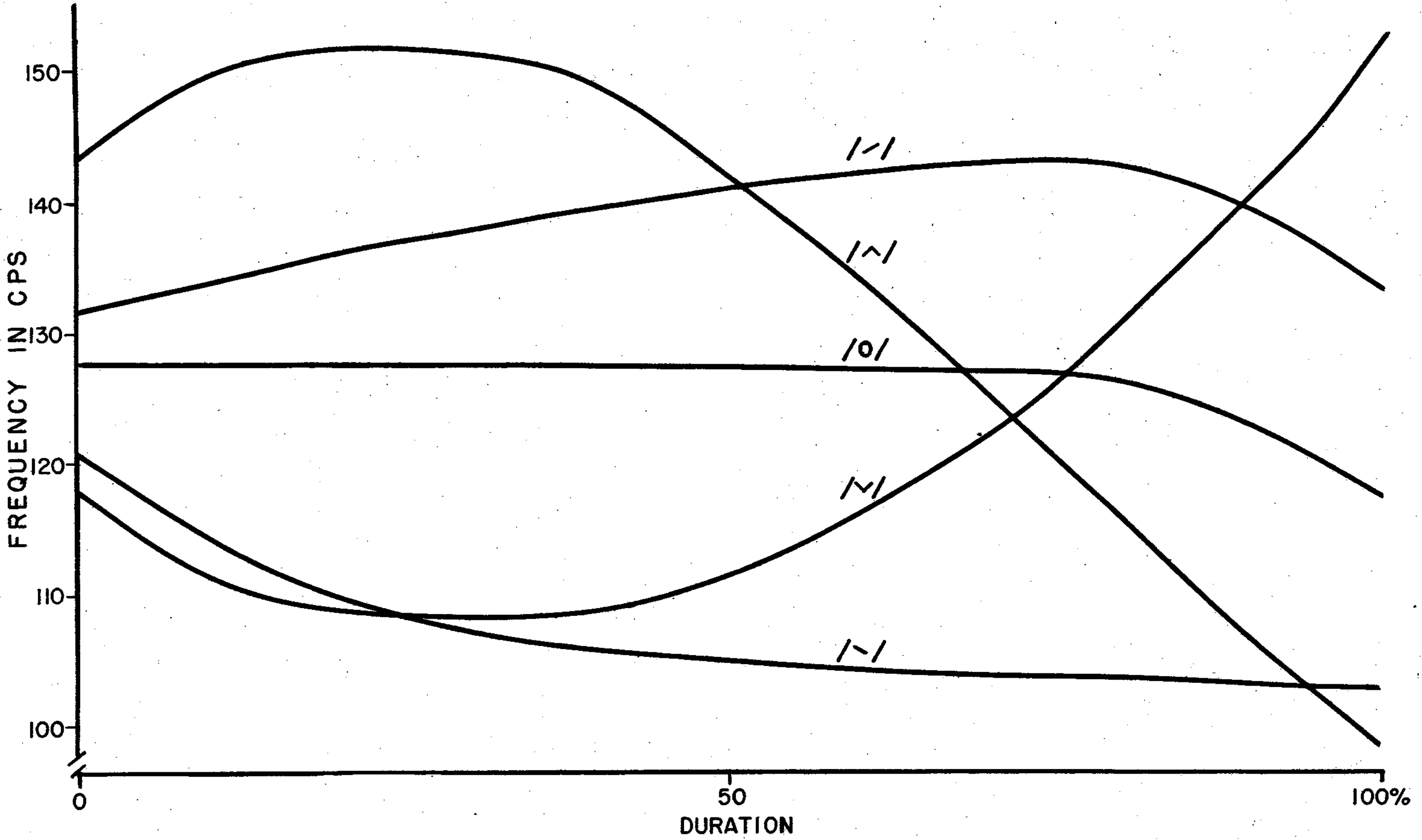


FIG. 3.6 TONES ON DOUBLE VOWELS

Figure 3.7 displays the fundamental frequency curves of these manifestations of the emphatic tone. They were arrived at in the same manner as for the phonemic tones. These curves are normalized in time and averaged in Figure 3.8.

B. Experiments on the Perception of Tonal Features

1. Monosyllabic Morphemes

To confirm a hypothesis that tones can be identified in isolated monosyllables without any help from tonal or verbal context, a perception test was prepared for each of four of the sets of tonally differentiated monosyllabic morphemes in Table 3.1. Recordings of syllables spoken by E. N. that were chosen for each test were multiplied by five and arranged in random order. All the tests were taken by eleven subjects, who were told to write their answers in Thai script. Glosses are to be found in Table 3.1.

Test 25: /naa nàa náa nâa nǎa/ x 5 = 25 test items.
Test 26: /lom lòm lóm lôm/ x 5 = 20 test items.
Test 27: /kʻaj kʻàj kʻáj kʻâj kʻǎj/ x 5 = 25 test items.
Test 28: /pʻææ pʻæ̀æ pʻǽæ/ x 5 = 15 test items.

Since most of the subjects scored 100% on all the tests, nothing much is gained by presenting four confusion matrices. Instead, the average scores are given in Table 3.4.

TABLE 3.4

Tonal Identification

Tests:	25	26	27	28
Av. %:	99.6	99.1	99.3	97.6

The slight drop in Test 28 is largely due to one person who had only 80% right.

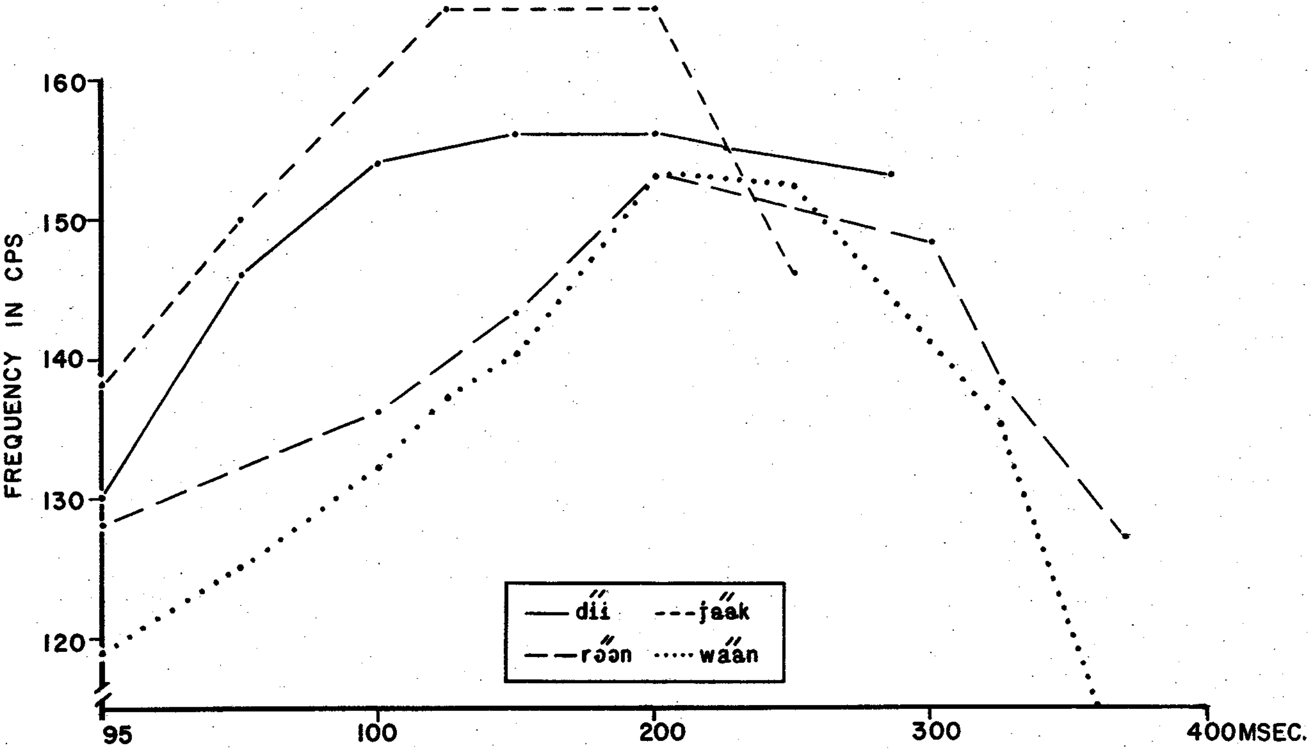


FIG. 3.7 EMPHATIC TONE

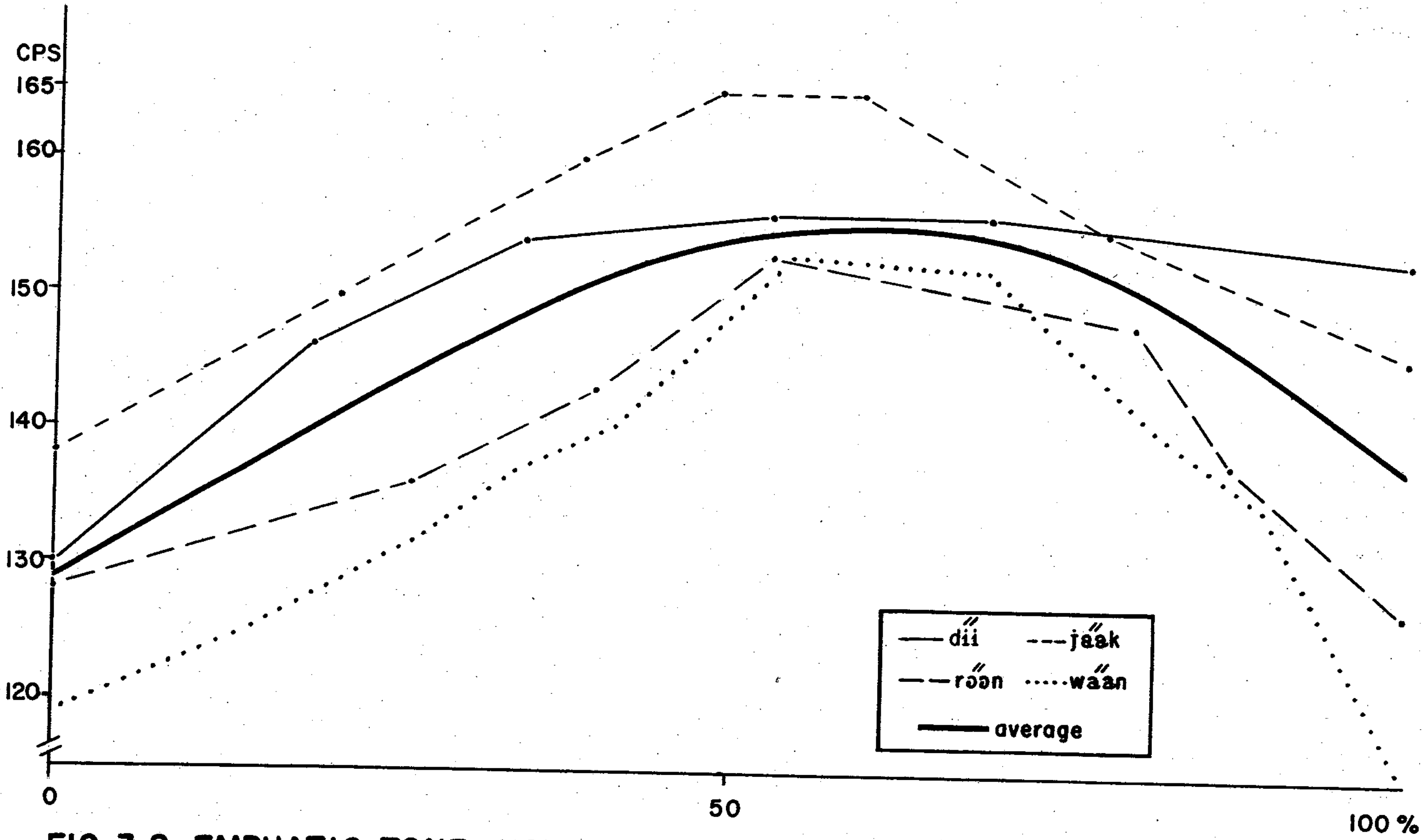


FIG. 3.8 EMPHATIC TONE: NORMALIZED CURVES AND AVERAGE CURVE.

2. Synthetic Tones

Experiments on the synthesis of Thai tones were run by means of the Intonator (Introd. C-1). The average tonal curves in Figures 3.5-6 were painted on the pitch-control tape to impose the corresponding pitch patterns on recordings of some of E. N.'s monosyllabic utterances in sets of five taken from Table 3.1. These artificially generated tones were presented to Thai subjects for identification. E. N. was no longer available to take the tests himself.

Test 29

The five average tonal curves that had emerged from analysis of syllables with single vowels were imposed on an utterance of the syllable /kʿaj/ (mid tone), which was chosen rather than any of the other four morphemes in the set, because it seemed desirable to work with a base that was as neutral as possible as to pitch. A syllable originally spoken on the mid tone was used in the next test for the same reason. The five resulting stimuli, taken from the output of the Intonator, were recorded on tape five times, making twenty-five test items, and arranged in random order for identification by six Thai subjects. Their responses are given in the form of a confusion matrix in Figure 3.9.

Test 30

The five average tonal curves for syllables with double vowels were synthesized on an utterance of /loo/ (mid tone). The five resulting stimuli were recorded on tape five times, making twenty-five test items, and arranged in random order for identification by six Thai subjects. Their responses are given in Figure 3.10.

At this point it was clear that the Thai tones had been successfully synthesized; however, the two small clusterings of confusions in Test 30 suggested that a further experiment be run to see whether the pitch curves carried enough information to override the effects of such concomitant features as variations in intensity, vowel quality and duration. Although a systematic examination of these features has not been conducted as a part of this study, it is obvious that they are present and audible. The two subjects who heard the intended /loó/ as /loo/ and another subject who always heard the intended /loô/ as /loó/, may have been influenced by the absence of features that normally accompany the high and falling tones. The experiment designed to investigate this matter involved imposing each of the synthetic tones on each member of a set of five tonally differentiated morphemes and obtaining identifications by Thai listeners. This was done with the Intonator, making use of its facility to substitute a new

	TONES PAINTED					
TONES HEARD	0	\	/	∧	∨	
0	30	3				
\		26				
/			30			
∧		1		30		
∨					30	
Totals	30	30	30	30	30	150

Intelligibility: 97.3%

FIG. 3.9 CONFUSION TABLE OF SYNTHETIC TONES ON /k'aj/ IN TEST 29.

	TONES PAINTED					
TONES HEARD	0	\	/	∧	∨	
0	30	1	8			
\		29				
/			22	5		
∧				25		
∨					30	
Totals	30	30	30	30	30	150

Intelligibility: 90.7%

FIG. 3.10 CONFUSION TABLE OF SYNTHETIC TONES ON /loo/ IN TEST 30.

fundamental frequency for the old one. While doing this, it retains the essential spectral and temporal characteristics of the original speech material, although there is a small degradation in quality that is inherent in analyzer-synthesizer work. In the present experiment, to the experimenter's satisfaction at least, concomitant features of the tones heard in the original recordings were clearly audible in the Intonator output.

Test 31

The five tonal curves for double vowels (Fig. 3.6) were imposed on each syllable of the set /naa nàa náa nâa nǎa/. The resulting twenty-five stimuli were recorded four times on tape, making 100 test items, and arranged in random order for identification by ten Thai subjects, including the six who had taken Tests 29 and 30. Their responses are shown in Figure 3.11. Twelve of the thirty confusions in the matrix were made by one person. In the light of the nearly perfect overall identifications achieved, it is interesting to note that the experimenter himself, in taking this test, was able to recognize nearly all instances of syllables originally spoken on the dynamic tones. He did this by associating audible allophonic variations in length and changes in the course of intensity with the falling and rising tones.

For Thai listeners, pitch is apparently the overpowering factor in the perception of Thai tones, no matter how striking certain concomitant features may be to the analyst.

C. Discussion of Results

1. Average Tonal Curves

The characteristic tonal curves presented in this study lend moderately good support to the usual auditory descriptions of Thai tones given by teachers of the language and textbooks.[9] Indeed, Kroll's impressionistic description of the pitch contours[10] accords rather well with the physical findings.

It is hard to tell whether auditory descriptions have not been influenced by the early instrumental findings of Cornelius B. Bradley,[11] especially since a slightly amended version of his tonal plot appears in a widely used dictionary.[12] Bradley's kymographic plots resemble the curves obtained spectrographically in a general way. Discrepancies are greatest in the high and falling tones. He shows the high tone dropping at the end to a point below the low tone. In the raw curves obtained for the present study, there is not even any marginal

	SYLLABLES SPOKEN	naa					nàa					náa					nâa					nǎa					
	TONES PAINTED	0	ˋ	ˊ	ˆ	ˇ	0	ˋ	ˊ	ˆ	ˇ	0	ˋ	ˊ	ˆ	ˇ	0	ˋ	ˊ	ˆ	ˇ	0	ˋ	ˊ	ˆ	ˇ	
SYLLABLES HEARD	naa	40		2			39		1			39	1	3			35	2				38		3			
	nàa		40		1			40		2	1		38		1			35		1		2	40				
	náa			38					39					37			5		39					37			
	nâa				39					38		1			39			2	1	39					40		
	nǎa					40	1				39		1			40		1			40					40	
	Totals	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	1000

Intelligibility: 97 %

FIG. 3.11 CONFUSION TABLE OF SYNTHETIC TONES ON 5 SYLLABLES IN TEST 31.

overlap of this sort. Bradley's falling tone shows no initial rise and starts its fall somewhat below the level of the mid tone. In the present study, all instances but one of the falling tone that have no initial rise, e.g. /sɨ̂ɨ/, start rather high in the speaker's voice range; the exception is /sɔ̂ŋ/. Aside from problems of instrumentation, Bradley's results were also limited by his small sampling of Thai speech, apparently little more than the /naa/ set of morphemes.

To the extent that Bradley's measurements do accord with the findings of the present study, it would be interesting to see whether auditory descriptions of the Thai tones published before 1911 are less strongly supported by acoustical measurements. One such description[13] seems to exist but was unobtainable.

Although considerably more than enough syllables were examined in the present study to arrive at the average tonal patterns of minimal utterances, a much larger number would be needed for a detailed allophonic study of the tones in all vocalic and consonantal environments. The present chapter could serve as a basis for such an investigation. Indeed, the frequency values of Tables 3.2-3, limited in this respect as they are, do at least contain some hints of what might be found. For example, the high tone on /lóm/ is unique in that it rises from the bottom of the speaker's voice range, but it turns out that the initial stretch from 83 to 120 cps coincides with the [l]. The variant of this tone before final stops and zero has already been mentioned (Fig. 3.5).

It should be noted that the emphatic tone is higher along most of its length than the high tone but is roughly similar to it in contour, even though it reaches its high point sooner than the high tone does. It is thus easy to understand, at least on phonetic grounds, why M. Haas considers it a modification of the high tone.[14]

2. Running Speech

The intersection of tones with sentence intonation is a complicating factor that remains to be explored. With this in mind, long passages of narrative and conversational material have been elicited from E. N. and K. S. and recorded on tape.

3. Perception

The results of Tests 25-28 make it evident that Thai subjects can identify the tones of monosyllabic utterances with a high degree of accuracy. These experimental data do not agree with R. B. Noss' impressionistic observation[15] that the mid and low tones are not distinguishable in isolation. Specifically, he says, /lom/ 'wind' and /lòm/ 'mudhole' can be distinguished only in an environment where relative pitch criteria are available, such as /pen ______/ 'It is . . . .'. No such difficulty was evident in Tests 25-28! Indeed, in Test 26 on the /lom/ set, a liberty was taken with the randomization. The first two items in the test order were made recordings of /lòm/ on the hypothesis that if Noss was right, the subjects would give random responses of /lom/ or /lòm/ or, at the very least, show a fair amount of confusion. This did not happen. That the mid and low tones are distinguishable in isolation is easy to understand, given the difference in shape between the two contours (Figs. 3.5-6).

An important extension of the experiments on the perception of tonal features would be a series of discrimination studies to determine the physical ranges of the tones and the perceptual crossover points between them. This can be undertaken with the successfully generated synthetic tones made for the present study as a point of departure.

Further work should be done also on phonetic features concomitant with the tones. An acoustic analysis will yield data of interest to problems of tonal perception in whispered speech and may also pave the way for experiments with synthetic speech in which pitch is held constant while variations are made along other parameters. The results of Test 31 seem to indicate quite clearly that pitch overrides the concomitant features, but it will be useful to see whether sufficient accentuation of these features can be made to override pitch in Thai subjects' judgments of tones. The experiment would, in a sense, be the converse of those comprising Test 31. Not only would the results be of linguistic interest but also of interest from the point of view of studies of auditory perception as well as that of engineering problems in speech transmission.

Notes

1. /kʻæ̂ŋ/ 'shin' was overlooked.
2. In /loolee/.
3. A kind of fish.
4. /kʻaaw/ 'odor' was overlooked.

5. The 100 cps tone used for making the calibration curve was included in each spectrogram.

6. This instrumental finding seems to be what Gedney predicted from listening; see "Indic Loanwords in Spoken Thai," p. 24.

7. For the glosses of the morphemes in the left-hand column, see Table 3.1.

8. The /naa/ set and the /kʻaj/ set.

9. E. g., B. O. Cartwright, *The Student's Manual of the Siamese Language* (Bangkok, 1915), pp. 29-31; M. R. Haas, *The Thai System of Writing* (Washington, D. C., 1956) pp. x-xi.

10. M. E. Kroll, "Suprasegmental Phonemes of Thai (Bangkok Dialect)" (unpub. master's thesis, Georgetown U., 1956), pp. 11-13.

11. "Graphic Analysis of the Tone-Accents of the Siamese Language," *JAOS*, XXXI (1911), 282-289. In a later publication, "On Plotting the Inflections of the Voice," *U. of Cal. Pubs. in Amer. Arch. and Ethnology*, XXII (1916), 195-218, he discusses distortions introduced by his earlier choice of scales and replots the falling and rising tones.

12. George B. McFarland, *Thai-English Dictionary* (Stanford, 1954 [orig. pub. 1941], p. x.

13. Caswell, "Treatise on the Tones of the Siamese Language" (Bangkok, ca. 1847, printed in *Siam Repository*, II, Bangkok, 1870).

14. M. Haas, "Techniques of Intensifying in Thai," *Word*, II (1946), 127-130.

15. "An Outline of Siamese Grammar" (unpub. Ph.D. dissertation, Yale U., 1954), Sec. 1.1.2.

SUMMARY AND OUTLOOK

Although a linguistic analysis of Standard Thai was not a primary goal of this study, a brief description of the phonology has been given as a framework for the acoustical measurements and experiments. Within the frame of the syllable, twenty consonants, nine vowels and five tones are established as the phonemes of Thai. Features of internal open juncture and sentence intonation are handled under syntagmatic phonology. What with its intersections with the distinctive tones, Thai sentence intonation is not easy to describe, but it is tentatively accounted for in terms of two pitch registers and three terminal junctures. Finally, there are the features of contrastive stress and emphatic tone. Of the total phonology outlined, the vowels, the tones and the emphatic tone have been examined acoustically and experimentally in the present study.

Formant frequencies of the nine vowels of Thai, occurring as single vowels, geminates and vocalic clusters, have been obtained from spectrograms. The emphasis here was on citation forms, but some measurements of vowel formants in a passage of narrative discourse have also been made. The prevalence of steady-state vowels in citation forms is a noteworthy characteristic of the language. Another is the wide range of variation of /a/, if it is said to subsume, as it is here, the second element of the three vocalic clusters. (Tables 1.1-4.)

In experiments on the perception of isolated vowels, Thai subjects had no trouble in identifying a speaker's vowels even when they were presented out of context. By means of similar identification tests, the subjects helped in arriving at a set of highly intelligible synthetic vowels specifiable in terms of two formants only. These two-formant synthetic vowels were based, initially, on an informant's average vowel productions. The synthetic versions were then revised several times to take the perceptual confusions into account. The synthetic vowels accepted as best by the listeners turned out to be very close in formant frequency values to the first two formants of W. N.'s original utterances, indicating that, as expected, most of the information on vowel quality is carried by the first two formants. (See Table 1.11.)

Data on the durations of citation forms of the vowels have been collected to furnish phonetic details on the phonemic contrast between single vowels and double vowels, with the latter including geminate vowels and the vocalic clusters /ia ɨa ua/. The length contrast was examined in connected discourse through measurements of vowels in a passage of running speech and in pairs of minimally distinguished morphemes embedded in a carrier sentence. (Tables 2.1-4.)

These data also show allophonic variations of length that are concomitant with the tones. At least for double vowels it can be said that vowels on the mid and low tones tend to be somewhat longer, on the high and falling tones, somewhat shorter, than on the rising tone. (Tables 2.5-7.)

Experiments on the perception of the distinction between single and geminate vowels yielded two major results: (1) There were distinct groupings of the single and geminate vowels, with a perceptual crossover zone between the two groups. (2) The groups were clearly distinguished by duration, thus strongly supporting the primacy of the quantitative feature over other features observed. (Fig. 2.5 and Table 2.8.)

A partial reservation to this conclusion must be made in the special case of isolated vowels synthesized on the Pattern Playback; two factors interfered with the discrimination of single and geminate vowels: (1) abrupt formant endings on geminate vowels, giving the impression of an unexpected glottal stop, and (2) instructions to the listeners which did not take into account their expectations as to tones. The matter of the glottal stop could not affect the results of the experiments on vowel duration, however, because the only open syllables used were on the high tone, and so a glottal stop was equally likely after both the single and double vowel of each pair. The matter of tones has no bearing on the general conclusion, since tones were kept constant in each experiment on vowel duration.

Measurements of the fundamental frequency patterns of many sets of tonally differentiated syllables have yielded average curves for the five phonemic tones on single vowels and on double vowels. An average curve was obtained for the emphatic tone as well. (Figs. 3.5, 6, 8.)

Experiments on the perception of the five tones had these major outcomes: (1) The tones are identifiable in isolated monosyllables. (2) Highly intelligible tones can be synthesized using the average pitch contours that emerged from measurements of real speech. (3) In the perception of tones, pitch convincingly overrides the effects of concomitant phonetic features observed in the utterances. (Table 3.4 and Figs. 3.9-11.)

Suggestions for further work stemming from certain aspects of the present study are made here.

An obvious extension of the study is an analysis of a large and systematic sampling of vowels and tones in connected speech. An important part of this task would be an examination of the vowels in all consonantal and tonal environments and the tones in all vocalic and consonantal environments.

The three major parts of this study provide an opportunity to make advances in our understanding of the influence of phoneme boundaries on the auditory discrimination of speech sounds, a topic that has loomed large in recent research on the perception of speech. Specifically, work at the Haskins Laboratories has thrown light on the relation between the ability of listeners to group speech sounds into phoneme categories by labeling the sounds and their ability to discriminate between such sounds. Experiments on the stop consonants of English have yielded rather sharp labeling functions and high peaks in discrimination at the phoneme boundaries. That is, subjects apparently interpreted these sounds only phonemically.[1] These results, taken together with somewhat different results from current work on English vowels,[2] have led to the following view of the perceptual process: a link between the perception of speech sounds and feedback from the articulatory movements that the hearer would use in producing these sounds, leads to categorical perception when the phonemic distinctions involve no intermediate articulations; in phonemic distinctions which rest on articulatory continua, perception will be less categorical, and differential discrimination ought then to be fairly constant throughout the range of variation.

The results and conclusions of the present study have given rise to three experiments designed to test the second half of this articulatory feedback theory. A test of the identifiability and discriminability of thirteen qualitative variants along the F1/F2 continuum on which Thai /ii ee ææ/ lie has been begun.[3] Another experiment in progress aims at doing the same thing for phonemic vowel length in Thai.[4] A similar experiment on the mid and high tones of Thai[5] yielded two results: (1) Difference in shape is a stronger cue than pitch height for this distinction. (2) The discrimination testing seemed to support, or at least not contradict, the hypothesis that there will be little or no sharpening of discrimination at this kind of phoneme boundary. The other two experiments, with fewer instrumental difficulties than in the tonal one, ought to provide a better test of the hypothesis.

An acoustic analysis of phonetic features concomitant with the tones is needed for a better understanding of tonal perception, especially in regard to the whispering of expressions that are distinguished by tones in normal speech. A separate investigation[6] has already shown that in whispered Thai, where there is no vocal cord vibration to produce a varying fundamental frequency, tones can be distinguished by features other than pitch, though there is a considerable and non-uniform reduction in the distinctions.

The effective factors in the perception of the emphatic tone could usefully be investigated. The experiments may be difficult, however, since the Thai do not appear to react to this emphatic feature as a separate category in the same neatly defined way as in the case of the phonemic tones.

The general subject of sentence intonation is a thorny one and is likely to be even more complicated in a language with phonemic tones. Research into the interaction of tones with sentence intonation in Thai is another possible extension of the work on tones described in this study.

Of some interest to linguistic and psychological theory would be some experiments on the perception of Thai vowels in terms of a reference system. That is, a test would be designed in which the listener had a chance to identify vowels with reference to the speaker's total vowel system as against a test in which no such opportunity was offered. The results could be compared with the results of similar experiments on other languages.

Finally, there is a vast amount of work yet to be done on consonantal features. The sampling of consonant-vowel formant transitions reported in this study represents a start in that direction.

Notes

1. A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The Discrimination of Speech Sounds within and across Phoneme Boundaries," J. of Exptl. Psych., LIV (1957), 358-368. For more recent thought and work, see A. M. Liberman, K. S. Harris, J. A. Kinney, and H. Lane, "The Discrimination of Relative Onset-Time of the Components of Certain Speech and Nonspeech Patterns," J. of Exptl. Psych., LXI (1961), 379-388.

2. D. B. Fry, A. M. Liberman, A. S. Abramson, and P. Eimas, in preparation.

3. In collaboration with Gloria F. Lysaught and David Crabb.

4. Jarvis Bastian and A. S. Abramson, in preparation.

5. A. S. Abramson, "Identification and Discrimination of Phonemic Tones," JASA, XXXIII (1961), 842 (A).

6. A. S. Abramson, "Vocoder Output and Whispered Speech in a Tone Language: Thai," JASA, XXXI (1959), 1568 (A). Also presented at the New York meeting of the Modern Language Association of America, December 28, 1958.

BIBLIOGRAPHY

The bibliography is a list of the sources cited in the text and not a compendium of all relevant literature.

Abramson, Arthur S. "Vocoder Output and Whispered Speech in a Tone Language: Thai," *Journal of the Acoustical Society of America*, XXXI (1959), 1568 (A).

________________. "Identification and Discrimination of Phonemic Tones," *Journal of the Acoustical Society of America*, XXXIII (1961), 842 (A).

Benedict, Paul K. "Tonal Systems in Southeast Asia," *Journal of the American Oriental Society*, LXVIII (1948), 184-191.

Borst, John M. "The Use of Spectrograms for Speech Analysis and Synthesis," *Journal of the Audio Engineering Society*, IV (1956), 14-23.

________________, and Cooper, Franklin S. "Speech Research Devices Based on a Channel Vocoder" [abstract], *Journal of the Acoustical Society of America*, XXIX (1957), 777.

Bradley, Cornelius B. "Graphic Analysis of the Tone-accents of the Siamese Language," *Journal of the American Oriental Society*, XXXI (1911), 282-289.

________________. "On Plotting the Inflections of the Voice," *U. of California Publications in American Archeology and Ethnology*, XII (1916), 195-218.

Cartwright, B. O. *The Student's Manual of the Siamese Language*. Bangkok: American Presbyterian Mission Press, 1915.

Caswell. "Treatise on the Tones of the Siamese Language," *Siam Repository*, II (Bangkok, 1870 [first pub. ca. 1847]).

Chang, Nien-Chuang T. (Mrs.). "Tones and Intonation of the Chengtu Dialect (Szechuan, China)," *Phonetica*, II (1958), 59-85.

Cooper, Franklin S. "Spectrum Analysis," *Journal of the Acoustical Society of America*, XXII (1950), 761-762.

Cooper, Franklin S. "Some Instrumental Aids to Research on Speech," *Report of the Fourth Annual Round Table Meeting on Linguistics and Language Teaching*. Washington, D. C.: Institute of Languages and Linguistics, Georgetown U., 1953, pp. 46-53.

Delattre, Pierre, Liberman, Alvin M., Cooper, Franklin S., and Gerstman, Louis J. "An Experimental Study of the Acoustic Determinants of Vowel Color; Observations on One- and Two-Formant Vowels Synthesized from Spectrographic Patterns," *Word*, VIII (1952), 195-210.

Delattre, Pierre. "Les indices acoustiques de la parole: Premier rapport," *Phonetica*, II (1958), 108-118, 226-251.

Dudley, Homer. "Remaking Speech," *Journal of the Acoustical Society of America*, XI (1939), 169-175.

Durand, Marguerite. *Voyelles longues et voyelles brèves: Essai sur la nature de la quantité vocalique*. (Collection linguistique publiée par la Société de Linguistique de Paris, XLIX) Paris: Librairie C. Klincksieck, 1946.

Egerod, Søren. *The Lungtu Dialect: A Descriptive and Historical Study of a South Chinese Idiom*. Copenhagen: Ejnar Munksgaard, Ltd., 1956.

Fischer-Jørgensen, Eli. "On the Definition of Phoneme Categories on a Distributional Basis," *Acta Linguistica*, VII (1952), 8-39.

Fowler, Murray and Israsena, Tasniya. *The Total Distribution of the Sounds of Siamese*. Madison: U. of Wisconsin Press, 1952.

Gedney, William J. "Indic Loanwords in Spoken Thai." Yale U.: unpub. Ph. D. dissertation, 1947.

Gillette, Josephine Anne. "Prosodic Features in Bangkok Thai." Georgetown U., unpub. master's thesis, 1955.

Greenberg, Joseph H. *Essays in Linguistics*. Chicago: U. of Chicago Press, 1957.

de Groot, A. W. "Phonologie und Phonetik als Funktionswissenschaften," *Travaux du Cercle linguistique de Prague*, IV (1931), 116-147.

Haas, Mary R. and Subhanka, Heng R. *Spoken Thai*. New York: Henry Holt, 1945.

Haas, Mary R. "Techniques of Intensifying in Thai," *Word*, II (1946), 127-130.

Haas, Mary R. "A Brief Description of Thai with Sample Texts," Outline for Types of Linguistic Structure. Berkeley: U. of California (Mimeo), 1956. 2d ed., pp. 1-27.

__________. Thai Vocabulary. Washington, D. C.: American Council of Learned Societies, 1955.

__________. The Thai System of Writing. Washington, D. C.: American Council of Learned Societies, 1956.

Haudricourt, André G. "Les phonèmes et le vocabulaire du thai commun," Journal asiatique, CCXXXVI (1948), 190-238.

Henderson, Eugénie. "Specimen: Thai," Le maître phonétique, LXIX (1940), 11-12.

__________. "Prosodies in Siamese: A Study in Synthesis," Asia Major, N.S.I. (1949), 189-215.

Hockett, Charles F. A Manual of Phonology. Baltimore: Indiana U. Pubs. in Anthropology and Linguistics, Memoir 11, 1955.

Joos, Martin. Acoustic Phonetics. (Language Monograph No. 23) Baltimore: Linguistic Society of America, 1948.

Kroll, Mary Elizabeth. "Suprasegmental Phonemes of Thai (Bangkok Dialect)." Georgetown U. unpub. master's thesis, 1956.

Kruatrachue, Foongfuang. "Thai and English: a Comparative Study of Phonology for Pedagogical Applications." Indiana U.: Ed. D. dissertation, 1960.

v. Laziczius, J. "Probleme der Phonologie," Ungarische Jahrbücher, XV (1935) cited by N. S. Trubetzkoy, Grundzüge der Phonologie. Prague: Travaux du Cercle linguistique de Prague, 1939.

Liberman, Alvin M., Harris, Katherine Safford, Hoffman, Howard S., and Griffith, Belver C. "The Discrimination of Speech Sounds within and across Phoneme Boundaries," Journal of Experimental Psychology, LIV (1957), 358-368.

__________, Harris, Katherine S., Kinney, Jo Ann, and Lane, H. "The Discrimination of Relative Onset-Time of the Components of Certain Speech and Nonspeech Patterns," Journal of Experimental Psychology, LXI (1961), 379-388.

Lotz, John. "Notes on Structural Analysis in Metrics," Helicon, IV (1942), 119-146.

McFarland, George Bradley. Thai-English Dictionary. Stanford, California: Stanford U. Press, 1954 (1st ed., Bangkok: 1941).

Moroney, M. J. *Facts from Figures*. Harmondsworth, Middlesex: Penguin Books, 1951.

Noss, Richard B. "An Outline of Siamese Grammar." Yale U.: unpub. Ph. D. dissertation, 1954.

Peterson, Gordon E. "The Phonetic Value of Vowels," *Language*, XXVII (1951), 541-553.

Pike, Eunice V. "Tonemic-Intonemic Correlation in Mazahua (Otomi)," *International Journal of American Linguistics*, XVII (1951), 37-41.

Pike, Kenneth. *Tone Languages*. Ann Arbor: U. of Michigan Press, 1948.

Pipping, Hugo. *Über die Theorie der Vocale*. (Acta Societatis Scientiarum Fennicae, Tom. XX, No. 11) Helsingforsiae: 1895.

————————. *Inledning till Studiet av de Nordiska Språkens Ljudlära*. Helsingfors: Soderstrom & Co., 1922.

Potter, Ralph K., Kopp, George A., and Green, Harriet C. *Visible Speech*. New York: Van Nostrand, 1947.

Trager, George L. "Siamese Phonemes: A Restatement" in *Studies Presented to Yuen Ren Chao on his Sixty-fifth Birthday*. Taipai, Taiwan: Academia Sinica, 1957.