ERRATA

Page 86, 3rd line from bottom, "the 0.1 or 0.2 level" should read "the 0.01 or 0.02 level".

Page 86, 6th and 9th lines from bottom, "the 0.5 level" should read "the 0.05 level".

Page 93, 2nd line from bottom, "aspeakers" should read "speakers".

Page 100, 9th line from bottom, "Pike (148:12-13)" should read "Pike (1948:12-13)".

Page 134, 6th line from bottom, ' prevocalic consonant." ' should read ' prevocalic consonant)." '.

Page 136, 3rd line from bottom, "the totals of" should read "the averages of".

Page 154, line 11, "(c) before a falling tone" should read "(c) after a falling tone".

Page 154, 8th line from bottom, "the falling tone" should read "the drop tone".

Page 156, line 5, column 5, " -- /ˎ/ " should read " -- /./ ".

Page 255, line 7, "Edinburg: Edinburg" should read "Edinburgh: Edinburgh".

THE ACOUSTIC AND PERCEPTUAL

NATURE OF TONE IN VIETNAMESE

VU THANH PHUONG

Thesis submitted for the Degree of Doctor of Philosophy of the Australian National University

June 1981

Except where otherwise acknowledged in the text, this thesis is the original work of the author.

Vũ Thanh Phương

ACKNOWLEDGMENTS

I am gratefully indebted to my successive supervisors, Dr David Bradley (now at the University of Melbourne), Dr Tim Shopen and Mr Phil Rose, whose valuable suggestions and insightful comments have guided my research and helped shape the thesis in its present form. I am very grateful also to Professor R. M. W. Dixon and the other members of the Department of Linguistics, The Faculties, the Australian National University, who not only have contributed to my training in this discipline, but also have provided a stimulating academic environment in which to work. Thanks are due specially to Dr John Bruce Millar, of the Computing Research Group, A.N.U., who initiated me to the technique of speech synthesis and allowed me access to his programs for synthetizing the tones used in perceptual tests described in Chapter Four, and to Mrs Pamela Johnston for typing the manuscript.

I wish to take this opportunity to express my thanks to the Governments of Vietnam and Australia for the arrangements that made this project possible, and to all those native speakers of Vietnamese who willingly served as informants or subjects in the endless recording and experimental sessions in Canberra, Sydney, Hanoi, Hue, Da Nang and Ho Chi Minh City.

Last but not least, my heartfelt gratitude goes to my family, friends and colleagues at home who have shared the burden of my absence and whose patience and sacrifices have been a constant source of inspiration for me.

Canberra

May 1981

ABSTRACT

This study presents a description of Vietnamese tones in the three major dialects - Northern, Central and Southern - in their various aspects and at different levels of analysis.

Acoustic phonetic materials were gathered from voice recordings of 34 informants, analyzed instrumentally and measured. The data obtained were treated statistically and normalized to yield typical values for the parameters of Fo, Intensity, Duration and Laryngealization, which serve to characterize all Vietnamese tones. The results established six non-stopped tones (level, falling, rising, drop, curve and broken) and two stopped tones (stopped rising and stopped drop) for Northern Vietnamese, which in this respect also represents Modern Standard Vietnamese. The Central and Southern dialects have only five non-stopped tones and two stopped tones, since there are no differences in the phonetic realizations of the standard curve and broken tones in these dialects. Four types of Fo contour (level, falling, rising and concave) and three levels of average Fo (high, mid and low) were established. Useful data on intensity, duration and laryngealization occurring on different tones were also obtained.

Analysis of the variation of tone parameters in their relationships with segmental and suprasegmental environments points to the complex interaction of various factors in the production of tone. The phonetic nature of Vietnamese tones can thus be understood to have acoustic properties related to the laryngeal mechanisms for the production of voice, with Fo contour, relative Fo level and voice quality as the basis for differentiating the tones .

Perceptual tests on cross-dialect tone recognition with natural and synthetic tones, involving 84 Northern, Central and Southern Vietnamese subjects, showed that natural tones of the three dialects in meaningful contexts were readily identified by all subjects, while isolated real speech tones and synthetic tones presented varying degrees of difficulty in recognition according to individual tones, dialects and subjects. Analysis of tone features involved in misperceptions suggested that contour features played a dominant role in tone identification.

A model of tone perception is proposed, including conversion processes and interpretation rules for translating the physical phonetic parameters into phonetic features (pitch targets and degrees of laryngealization) then into phonological features (contour, concave, high, low, falling, rising, creaky, stopped). The model is supposed to reflect the dynamic process of tone perception in which the phonetic properties of tones are perceived in terms of the native speakers' phonological structures.

TABLE OF CONTENTS

Contents Page

Contents Page

Contents Page

LIST OF TABLES

Tables Page

Tables Page

LIST OF FIGURES

Figures | Page

Figures		Page

LIST OF ABBREVIATIONS

A, B, C...	Terms of a formula or a rule
A(B)	B is a possible alternative to A
$\begin{Bmatrix} A \\ B \end{Bmatrix}$	Either A or B
C	Consonant (or glide); Central
(C1), (C2)...	Conversion formulae
CF	Central (Vietnamese) Female (Informant)
CM	Central (Vietnamese) Male (Informant)
CO	Consonant onset
cont	Continuant
cor	Coronal
Cor	Correct (responses)
Cr	Creaky
CV	Central Vietnamese
D	Duration
dB	Decibel
E	Endpoint
F	Female; Frequency
$\bar{F}$	Mean Fo level
F1, F2...	Formant 1, Formant 2...
Fa	Falling
FD	Frequency Differential
FD($\bar{F}$)	Frequency Differential relative to the mean $\bar{F}$
Fi	Individual Fo value
Fo	Fundamental frequency

Abbreviations (continued)

Hi	High
Hz	Hertz
I, Int	Intensity
I1, I2...	Intensity measurement points
Itg	Integer
L, Lar, Laryng	Laryngealization, Laryngealized
Lo	Low
M	Male; Midpoint
max	Highest value
min	Lowest value
N	Northern
n	Number of occurrences
(N1), (N2)...	Normalization formulae
NF	Northern (Vietnamese) Female (Informant)
NM	Northern (Vietnamese) Male (Informant)
ns	Non-significant
NV	Northern Vietnamese
O	Onset
p	Probability; page
P1, P2...	Fo measurement points
PTR	Phonological tone rule
PU	Pitch unit
R	Range; Responses
RC	Redundancy convention
Ri, Ris	Rising

Abbreviations (concluded)

S	Southern; Standard deviation (in formulae)
s	Stopped (in notations)
S^2	Variance
SD	Standard deviation (in tables)
SF	Southern (Vietnamese) Female (Informant)
SM	Southern (Vietnamese) Male (Informant)
son	Sonorant
St	Stopped
Stim	Stimuli
Subj	Subjects
SV	Southern Vietnamese
T	Tone
t	t value calculated for t-tests
TI	Tone interpretation rule
TRC	Tone redundancy convention
V	Vowel; Vietnamese
VN	Standard Vietnamese
$\bar{x}$	Mean value
χ^2	Chi-square value calculated for χ^2-tests

LIST OF SYMBOLS

Phonological notation for tone

/-/	Level tone	e.g.	/ma̅/	'ghost'
/\`/	Falling tone		/mà/	'but'
/´/	Rising tone		/má/	'cheek'
/´s/	Stopped rising tone		/mát/	'cool'
/.../	Drop tone		/mạ/	'rice seedling'
/.s/	Stopped drop tone		/mạt/	'wretched'
/ʾ/	Curve tone		/mả/	'tomb'
/~/	Broken tone		/mã/	'horse' (as chessman)

Phonetic notation for tone (examples)

˧	3	Level, mid pitch
˧˥	35	Rising, high pitch
˨˦	24	Rising, mid pitch
˨˩s	21s	Falling, low pitch, stopped
˨˩̰	21̰	Falling, low pitch, creaky (laryngealized)
˨˩˨ (̤)	212 (̤)	Concave (falling-rising), low pitch, (optional breathy voice)
		etc.

$	Syllable boundary
→	Become (in transformational rule); imply (in redundancy convention)
-:	Perceived, interpreted or reinterpreted as
>	(Historically) change to; greater than
<	(Historically) come from; less than
*	(Before a phonological form) not occurring (Before an historical form) reconstructed
∅	Zero, null
Σ	The sum of

<u>Note</u> Other usual phonetic symbols follow the principles of the International Phonetic Association 1949.
Other common mathematical symbols have their usual values.

CHAPTER ONE

INTRODUCTION

1.1 *AIM AND SCOPE OF STUDY*

This study is intended to provide a description of Vietnamese tones in their various aspects and at different levels of analysis.

My purpose is to discover the acoustic, articulatory and perceptual features of Vietnamese tones as produced by the speakers of the three major dialects (Northern, Central and Southern Vietnamese, henceforward NV, CV and SV, to be defined in 1.4 below) and as perceived both within dialect groups and across dialect boundaries.

Vietnamese tones have been described in a number of grammars and phonological studies. In spite of many differences, most of these studies agree on a few major points: the number of tones, certain auditory and acoustic characteristics that differentiate them (although descriptive labels vary a great deal; see Earle 1975:42), the association of tone with the syllable and its function at the lexical and semantic levels.

To mention some of the better known works, impressionistic descriptions of NV tones can be found in Thompson 1965, Nguyen Dinh Hoa 1967, Hoang Tue and Hoang Minh 1975 and Doan Thien Thuat 1977, of SV tones in Thompson 1959 and Henderson 1961, and of CV tones in Emeneau 1951. Contrastive descriptions of NV and SV tones are given in Jones and Huynh Sanh Thong 1957 and Nguyen Dang Liem 1970. An impressionistic survey of tones in six dialectal areas - Hanoi, Vinh, Hue, Da Nang, Saigon (now Ho Chi Minh City) and Tra Vinh - can be found in Thompson 1965.

Descriptions based on instrumental analysis are given in Le Van Ly 1948, Andreev and Gordina 1957, Nguyen Ham Duong 1962, Han 1969, Han and Kim 1972 and Earle 1975, for NV tones, and in Tran Huong Mai 1969 for SV tones.

The only perceptual study of Vietnamese tones that I know of is by Miller 1961 on recognition of NV tones in whispered speech. The historical development of tone in Vietnamese has been discussed in Haudricourt 1954, 1961 and 1972, Thompson 1969 and Matisoff 1973. A reconstruction of Proto-Viet-Muong tones has been attempted by Thompson 1976 in a study of Proto-Viet-Muong phonology.

It appears from the above list that the tones of NV, regarded as a prestige dialect closest to the standard national language, have been studied more extensively and CV tones are the least described of all. Moreover, with a few exceptions, most of these studies tend to describe only standard tone forms without attempting to show how these forms can vary in different environments, and little or no attention has been paid to variation among dialects or variation among individuals within those dialects. This study will attempt to fill some of these gaps.

1.2 DESCRIPTIVE FRAMEWORK AND DEFINITION OF TERMS

I posit three levels of representation for tonal description.

At the physical phonetic level, tones are viewed as syllable-specific properties of the laryngeal voice produced in the phonatory process, as different from the formant structures, transitions or aperiodic noise which characterize segments. These properties can be extracted from acoustic data and described by means of 'physical phonetic parameters,' which in my framework represent the normalized values, in numerical scales, of Fo (fundamental frequency), I (relative intensity), D (relative duration) and L (laryngealization). The first three parameters have their usual meanings in acoustic phonetics. In my study, laryngealization is used as a cover term for some tone-specific phonation types which suggest "a tendency to constriction at the laryngeal level" (Laver 1980). This includes varying degrees of creaky voice, breathy voice and glottal closure which occur as distinctive or redundant features of some tones, and therefore has a wider meaning than that used in some other works (e.g. Ladefoged 1971, 1975 and Lehiste 1970, where laryngealization is equivalent to creaky voice.)

Within the same physical phonetic level, the phonatory process for tone production will be referred to in terms of laryngeal and aerodynamic factors which underlie the articulatory gestures.

At the systematic phonetic level, tones will be described in terms of n-ary phonetic features, namely pitch and laryngealization, using the five levels of pitch in Chao's system (1930) and four degrees of laryngealization as proposed in my scheme (Ch. Two). I take this level to represent the auditory and articulatory reality of speech sounds, usually recorded in systematic phonetic transcriptions with the details and variations which trained phoneticians can hear and observe directly without instruments.

At the phonological level, which in my view represents the psychological reality of speech in the native speakers' system, tones will be characterized by binary phonological features of the type proposed in Wang 1967, with some modifications. I adopt the following phonological features for the description of Vietnamese tones: 'contour,' 'concave,' 'high,' 'low.' 'falling,' 'rising,' 'creaky,' and 'stopped.' The reasons for this choice will be explained in 1.5.4.3 infra.

The following example will show, briefly, how the tones will be represented at different levels.

The SV curve tone has a concave contour, a mid level of average Fo, and no characteristic laryngealization. The mean Fo values in Hertz, derived from acoustic measurements at six time points, for a group of nine SV informants and an individual female informant SF1, are respectively

(SV group)	173	162	149	175	219	224
(SF1)	306	286	256	275	315	317

Normalization procedures proposed in Ch. Two will bring these figures into more comparable forms with the parameter Fo Differential, in percent, relative to the mean $\bar{F}$ (in this case 183 Hz for the SV group and 291 Hz for Informant SF1) :

(SV group)	-5	-12	-18	-4	19	22
(SF1)	5	-1	-12	-5	8	8

These figures give us a better idea of the typical shape of this tone by the group as a whole and by Informant SF1. The negative values represent Fo levels below the mean $\bar{F}$.

At the systematic phonetic level, either determined auditorily or by means of conversion processes proposed in Ch. Five, this tone will be represented in its standard form for SV as

$$\begin{bmatrix} 214 \text{ Pitch} \\ 0 \text{ Laryng} \end{bmatrix} \quad \text{or more briefly } [214]$$

and for Informant SF1 as

$$\begin{bmatrix} 313 \text{ Pitch} \\ 0 \text{ Laryng} \end{bmatrix} \quad \text{or } [313]$$

Possible variants for this tone, depending on individual speakers or phonetic environment, are [215], [324], etc. These details and variations concerning the onset, midpoint and endpoint pitch targets in n-ary values are auditorily perceptible, but will become irrelevant at the phonological level when this tone could be represented as

$$\begin{bmatrix} \text{T} \\ (+ \text{ contour}) \\ (+ \text{ concave}) \\ - \text{ high} \\ - \text{ low} \\ + \text{ falling} \\ + \text{ rising} \\ - \text{ creaky} \end{bmatrix}$$

The choice of these features will be explained in 1.5.4.3 infra and the way they could be derived from the phonetic representations will be described in Ch. Five. Briefly, the features [- high, -low] characterize the tone as having mid average pitch in contrast to a higher tone or a lower tone. The features [+ falling] and [+ rising] refer respectively to the initial downward slope and the subsequent upward slope of its pitch contour. The feature [- creaky] reflects the lack of characteristic laryngealization. The features in parentheses can be predicted by redundancy conventions. Thus this SV curve tone can be conveniently described as a mid-concave or mid-falling-rising tone.

The basic assumptions underlying the framework proposed are that the three levels of representation reflect the complex nature of tonal facts existing as acoustic, articulatory and psychological realities and that these realities can be observed, analyzed and shown to be inter-related within an integrated linguistic system.

The three levels will be formally described and related to each other by various rules and processes presented in Ch. Five. As an addition to current theory, I shall propose a model of tone perception with some forms of tone interpretation rules to represent the perceptual, as distinct from the generative, process. I have not come across any such model, but only a few hints from Sampson 1970 that phonological rules could be made reversible to capture the derivation, performed by the hearer, of underlying representations from surface representations.

1.3 GENERAL PROCEDURE AND PLAN OF STUDY

To determine the phonetic characteristics of the tones, I shall derive acoustic phonetic data from measurements of mingograms and spectrograms of NV, CV and SV informants' voices and treat them statistically to obtain physical phonetic parameters in normalized values. I shall refer to underlying physiological mechanisms for tone production in terms of laryngeal and aerodynamic factors that can be related to the above data and those supplied in other studies.

The perceptual study of tone will be based on perceptual tests involving both natural and synthetic tones, administered to native speakers of the three major dialects.

In accordance with this procedure, my study will be organized as follows.

In the rest of this chapter, I shall provide some background information on Vietnamese and its dialects and some basic facts of Vietnamese phonology.

Chapter Two deals with the acoustic aspects of Vietnamese tones with emphasis on their variation across the major dialects, and the way they could be related and compared through normalization procedures.

In Chapter Three, I shall enquire into the physiological and physical nature of tone by examining the various intrinsic factors that exert influence on Fo variation, and referring them to the phonation process.

Chapter Four presents the results of cross-dialect perceptual tests and an analysis of various factors involved in tone perception.

In Chapter Five, I shall propose a model of tone perception that is presumed to underlie the perceptual mechanism.

1.4 VIETNAMESE AND ITS DIALECTS

The Vietnamese language, spoken by about fifty million people, is remarkably homogeneous in its basic structures. Although variations exist in many respects, they are often overlooked or ignored by linguists who are more interested in describing the idealized standard forms.

Opinion differs as to the number of dialects and the degree of variation (cf Maspero 1912, Thompson 1965, Vuong Loc 1975, Nguyen Van Tu 1976, among others). I share the view of most linguists in describing Vietnamese as having three major dialects, the Northern, Central and Southern. Each major dialect comprises several local dialects that differ from each other in some respects, but data on these differences are not readily available. In this study, I shall refer to NV, CV and SV as the standard forms of those dialects represented by the better known speech of Hanoi, Hue and Ho Chi Minh City and their surrounding areas respectively. Between them, there is little variation in syntax, a common core of over 90% of the basic vocabulary, and most variations occur in the realm of phonology and phonetics.This enables native speakers of Vietnamese readily to recognize a Northern, Central or Southern accent while encountering little difficulty in mutual understanding (perhaps with the exception of some remote local dialects.)

Some of the major phonological differences between NV, CV and SV can be found in Vuong Loc 1975 and in Thompson 1965, who noted "an overall picture of gradually changing patterns from north to south." It does not follow, however, that the NV and SV dialects are farther apart because of geographical distance. There are, for example, more phonetic similarity between NV and SV tones than between these and the CV system.

A sketch map of Vietnam's dialectal structure is presented in Figure 1.1 (next page). As we can see from this map, Hanoi and Ho Chi Minh City are two important urban centres from where most

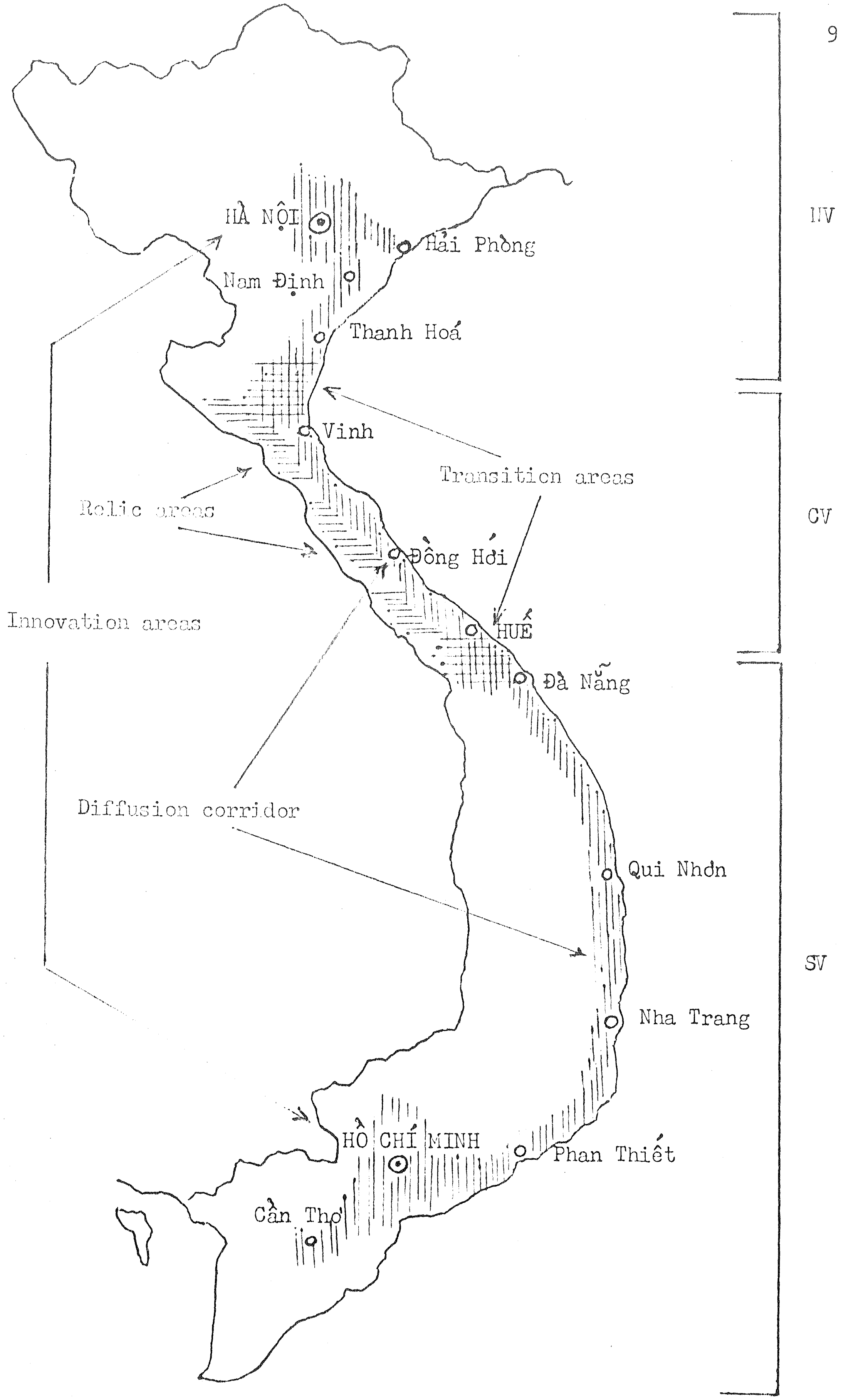


FIGURE 1.1 Viet Nam's Dialectal Structure

linguistic innovations are supposed to originate and spread along diffusion corridors (vertical hatching) passing through a string of coastal towns. The two transition areas (cross hatching) represent places where one can find some mixed features of NV and CV (the region between Thanh Hoa and Vinh) and of CV and SV (the region between Hue and Da Nang). Relic areas (horizontal hatching) are located in various CV provinces, especially in rural districts, where local dialects have retained many elements of earlier Vietnamese in terms of vocabulary and pronunciation.

The concurrence of various factors - common core of vocabulary, common grammar structures, gradual changes of phonological features between any two dialects, plus the modern tendency toward standardization through the media and education - has contributed to mutual intelligibility between dialects and greater homogeneity of the national language.

1.5 A SKETCH OF VIETNAMESE PHONOLOGY

This very brief sketch is given with a dual purpose in mind: to provide the reader with some basic understanding of the Vietnamese sound system and to illustrate the more general phonological framework to which subsequent descriptions of tonal phenomena could be related.

1.5.1 The Vowel System

The vowel system is given in Table 1.1 (next page). Redundant (predictable) features for vowels are not given in the table, as they can all be predicted by a Redundancy Convention of the form

TABLE 1.1 Phonological System of Vietnamese Vowels

Class Labels → / ↓ Features → ↓		FRONT [- back] [- round] [- tense]	FRONT [+ tense]	CENTRAL [+ back] [- round] [- tense]	CENTRAL [+ tense]	BACK [+ back] [+ round] [- tense]	BACK [+tense]
HIGH	[+ high] [- low]	i (i,y)*	iə**(iê,yê, ia,ya)	ɯ (ư)	ɯə (ươ,ưa)	u (u)	uə (uô,ua)
MID	[- high] [- low]	e (ê)		ɤ̆*** (â)	ɤ (ơ)	o (ô)	
LOW	[- high] [+ low]	ɛ̆ (a)	ɛ (e)	ă (ă,a)	a (a)	ɔ̆ (o)	ɔ (o, oo)

* Forms in parentheses are those of the official orthography.

** Phonological tense high vowels are phonetically diphthongized (though not in all local dialects), hence this notation.

*** Those lax vowels are also short when thus marked.

A → B which automatically assigns the redundant features to a vowel which is always [+ syl(labic)] in Vietnamese.

$$\text{(RC1)} \quad [+\ \text{syl}] \rightarrow \begin{bmatrix} -\ \text{cons(onantal)} \\ +\ \text{voice} \\ +\ \text{son(orant)} \\ +\ \text{cont(inuant)} \end{bmatrix}$$

In subsequent notations, unless otherwise required for greater clarity, features will be given in their abbreviated forms, and [+ syl] or [V] will stand for a vowel.

1.5.2 The System of Consonants and Glides

Some phonologists (e.g. Maspero 1912, Hoang Tue and Hoang Minh 1975, Doan Thien Thuat 1977, Nguyen Phan Canh 1978, among others) would recognize two separate systems of Vietnamese consonants, the initial and the final ones, and some would even recognize the medial /w/ as a separate class by itself. I would resolve whatever is different between them in function of their position in the syllable structure by a special feature such as [- release] in a phonological rule, which can aptly describe the main difference between initial and final stops, while giving them the same description in terms of phonological features and including them in one and the same system.

The system is presented in Table 1.2 (next page). The abbreviations 'cor', 'asp', 'son' and 'cont' in the table stand for 'coronal,' 'aspirated,' 'sonorant' and 'continuant' respectively. Note that in subsequent notations, [- syl] or [C] will stand for a consonant or a glide, and

TABLE 1.2 Phonological System of Vietnamese Consonants and Glides

Class Labels → / ↓ Features → ↓				LABIAL	ALVEOLAR	RETROFLEX	PALATAL	VELAR	GLOTTAL
				[- high] [- back] [- cor]	[- high] [- back] [+ cor]	[+ high] [- back] [+ cor]	[+ high] [- back] [- cor]	[+ high] [+ back] [- cor]	[- high] [+ back] [- cor]
OBSTRUENT [- son]	(ORAL) STOP	[- cont]	[- voice]	p* (p)**	t (t)	ʈ (tr)	c (ch)	k (k, c, q, -ch)	
			[+ asp]		t^h (th)				
			[+ voice]	b (b)	d (đ)				
	FRICATIVE	[+ cont]	[- voice]	f (ph)	s (x)	ʂ (s)		x (kh)	h*** (h)
			[+ voice]	v (v)	z (d, gi)	ʐ (r)		ɣ (g, gh)	
SONORANT [+ son]	GLIDE & LIQUID***	[+ cont]	[+ voice]		l (l)		j* (i, y)	w (u, o)	
	NASAL (STOP)	[- cont]	[+ voice]	m (m)	n (n)		ɲ (nh)	ŋ (ng, ngh, -nh)	

*/p/ and /j/ occur as syllable codas only, in underlying forms.

** Forms in parentheses are those of the official orthography.

*** The glides also include /h/, which is non-sonorant; the liquids also include [ɽ] in some dialects as a surface realization of /ʐ/; /w/ is a labio-velar glide; the sonorants also include vowels.

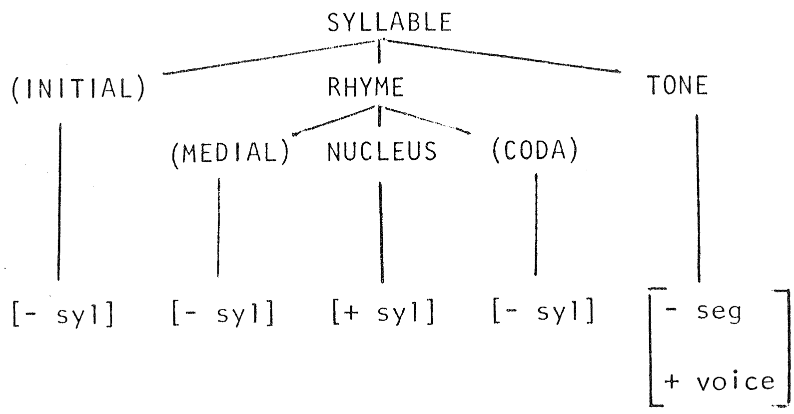
$$\begin{bmatrix} C \\ -\ \text{cont} \\ -\ \text{son} \\ -\ \text{voice} \end{bmatrix}$$

will represent a voiceless stop, which is of direct relevance to tonal distribution in Vietnamese.

1.5.3 The Vietnamese Syllable Structure

Writers of Vietnamese grammars and phonology have always viewed the syllable as a phonological unit. Though the syllable has not been given formal status in the standard generative theory, its phonetic and phonological reality has been shown by various authors such as Wang 1967, Kratochvil 1968, Fudge 1969, Sampson 1970, Hooper 1976, Bell and Hooper 1978, Lehiste 1978, Fujimura and Lovins 1978, among others.

The Vietnamese syllable can be described as having the following structure in a non-linear tree-diagram, in which elements in parentheses can be zero, and the features [- seg(ment), + voice] mark the tone. By convention, tonal features can be marked thus, or alternatively, with [T], whenever confusion with segmental features may arise.



This structure is advocated by most writers on Vietnamese phonology, although in different formal representations. Note that I have not used the term 'Final,' to avoid possible confusion, because in most studies of Vietnamese phonology, this term means 'final consonant,' while in most descriptions of Chinese phonology, it is the equivalent of 'rhyme' in my diagram.

This structure implies that in a Vietnamese syllable, the vowel nucleus and tone are the only obligatory elements. There are a number of phonotactic constraints which will establish the following.

(a) The system of initials comprises all consonants and glides in Table 1.2 except /p/ and /j/. Initial /p/ occurs only in a few loan words. Initial /j/ occurs only as a surface variant of /z/ in some dialects.

(b) The medial can only be /w/. It cannot occur if initial /w/ is present, and occurs after labial initials only in a few loan words from French, e.g. /mwa/ 'me' (< Fr moi), /vwan/ 'veil' (< Fr voile).

(c) The system of codas (final consonants and glides) comprises only eight underlying phonemes and four allophonic variants as given below.

p	t	(c)	k	(k^w)
m	n	(ɲ)	ŋ	($ŋ^w$)
		j	w	

The variants can be derived by rules, e.g. final velars are labialized after the back rounded lax vowels. Thus /ók/ 'periwinkle'

→ /ók^w/, /ɔ̄ŋ/ 'bee' → /ɔ̄ŋ^w/, etc.

(d) The short vowels /ɛ̆/, /ă/, /ɤ̆/ and /ɔ̆/ occur only in closed syllables, such as /tắk/ 'stopped,' /kɤ̆̄m/ 'dumb,' /mɔ̆́k/ 'hook,' and /ɛ̆̄ŋ/ 'brother,' but not in */tă/ or */kɤ̆/.

(e) Final /j/ does not occur after front vowels and final /w/ does not occur after back rounded vowels. The following combinations occur:

/iw/ /iəw/	/ɯw/ /ɯəw/ /ɯj/ /ɯəj/	/uj/ /uəj/
/ew/	/ɤ̆w/ /ɤw/ /ɤ̆j/ /ɤj/	/oj/
/ɛw/	/ăw/ /aw/ /ăj/ /aj/	/ɔj/

This points to a regular dissimilatory pattern and the behaviour of central vowels (some of them are back unrounded phonetically) as a group, before final glides.

1.5.4 The Tonal System

Vietnamese is a tonal language in the sense that it has "lexically significant, contrastive, but relative pitch on each syllable," as defined in Pike (1948:3).

Table 1.3 (next page) gives the reader a general idea of the system and its variation across dialects. The following comments are necessary.

1.5.4.1 <u>Tone Names</u> Most authors writing on Vietnamese agree that modern standard Vietnamese has six underlying tones and two conditioned variants that occur only in syllables with voiceless final stops, but few have discussed the question at length and some tend to give the variants very little attention in their description. In my study, I shall use the term 'phonological tones' when referring

TABLE 1.3 The Tones of Vietnamese

Number *	1	2	3	3B*	4	4B*	5	6
Vietnamese Names *	ngang	huyền	sắc	sắc (tắc)	nặng	nặng (tắc)	hỏi	ngã
English Labels *	level tone	falling tone	rising tone	stopped rising tone	drop tone	stopped drop tone	curve tone	broken tone
Phonological Notations *	/-/	/\/	/´/	/´s/**	/./	/.s/**	/ʼ/	/~/
Phonetic Notations *								
NV	[33]	[21]	[35]	[45s]**	[21̰]**	[21s]**	[212]	[32̰5]**
CV	[55]	[42]	[24]	[34s]**	[31]	[31s]**	[312̰]**	
SV	[33]	[21]	[35]	[35s]**	[212]	[21s]**	[214]	
Examples	/hāj/ 'two'	/hàj/ 'slipper'	/háj/ 'to pick (fruit)'	/hát/ 'to sing'	/hạj/ 'harm'	/hạt/ 'grain'	/hảj/ 'sea' (in compounds only)	/hãj/ 'scared' (in compounds only)

* See comments in the text.

** s represents the syllable-final voiceless stop which conditions the occurrence of the tone;

~ marks the laryngealization characteristic of the tone.

to the six underlying tones, 'phonetic tones' when speaking of their phonetic realizations at surface level, which number seven or eight according to dialects, and 'stopped tones' or 'stopped variants' interchangeably for the conditioned variants occurring at surface level only. In phonological and phonetic notations, this conditioning factor is included by marking the variants with a small s, as there is reason to believe (and I shall try to show later) that this factor underlies both the occurrence of the variant forms and their recognition as such by the hearer.

The Vietnamese names do not include special reference for these stopped variants, although the word 'nhập' ('entering') is sometimes used. I prefer to qualify them by the conditioning factor mentioned above. The English names are taken from Han 1969 and preferred to other labels for a number of reasons: they are short and suggestive of the basic contours of each tone; they reflect the fact that across dialects the tones tend to retain their basic contours while varying in relative pitch. This fact in turn may be reflecting a typological characteristic of Vietnamese tones as having primacy of contour over relative pitch. This is another proposition I shall examine later.

1.5.4.2 Number of Tones and Tone Numbering The number of tones in NV and SV presents no problem, as all authors agree that NV has six phonological tones and two stopped variants, while SV has one less because the curve and broken tones have historically merged. The picture is more complex with CV tones. Although the standard dialect in Hue and most of the CV area has five tones (Maspero 1912, Thompson 1965), there are reports of a six-tone system at Vinh (Emeneau 1951) and also a four-tone system (Duong Thanh Binh 1971).

It appears, from my recorded data and personal contacts with CV speakers, that the six-tone system only occurs at some places in the 'transitional area' between CV and NV (see map on Figure 1.1). The four-tone system occurs only in some local CV dialects in rural districts. It is therefore adequate to describe standard CV as having five tones, as in SV, i.e. without the broken tone. This is supported by the acoustic data and auditory impressions I got from informants coming from various CV provinces.

For the phonological notations, different systems of numbering have been used. Therefore, I have decided against such a system, because it may cause confusion for the readers who have been used to another system. Instead, I have settled for the diacritic signs used in the official Vietnamese orthography, with the exception that a macron is used for the level tone which is left unmarked in usual spelling. This will favour the reader who knows some Vietnamese, and I hope others will get familiarized with them easily.

Though I am not going to use it, the numbering system which underlies the presentation order in my description deserves some explanation. This order is chosen for various reasons: one of them is relative stability and another is relative frequency of occurrence. The first three tones, the level tone /-/, the falling tone /ˋ/ and the rising tone /ˊ/, practically retain their basic contours across dialects and account for a total of nearly 70% of occurrences in the basic vocabulary and nearly 80% of occurrences in speech. The tone numbered 6, the broken tone /~/, occurs almost exclusively in NV and have merged with the curve tone /ʼ/ in CV and SV. Moreover, the curve and broken tones, taken together, account for only 15%

of occurrences in the basic vocabulary and about 10% in speech (see Table 1.4, next page).

1.5.4.3 <u>Tone Features</u> As noted in 1.2 (p 4-5), beside the physical phonetic parameters in numerical scales, I have chosen two systems of features to represent Vietnamese tones at two other levels. The phonetic features of 'Pitch' and 'Laryngealization' represent tones at the systematic phonetic level, where details concerning pitch levels at onset, midpoint and endpoint, and several degrees of laryngealization are necessary to distinguish between variants of the same phonological tones. These features correspond to the notation system proposed in Chao 1930, with five levels of pitch and one, two or three pitch targets depending on whether the tone is level, with unidirectional or bidirectional contour. To mark laryngealization, the diacritic signs [~] (creaky voice) and [..] (breathy voice), after Ladefoged 1971 and Laver 1980, are also used under the pitch number or Chao's tone letters. The standard phonetic tones of NV, CV and SV can be thus represented in short phonetic notations as in Table 1.3 above (p 17). The values given were first obtained impressionistically, then readjusted (in a few cases only) when they could be derived from acoustic data according to conversion formulae proposed in Ch. Five (section 5.2).

At the phonological level, the features 'contour,' 'concave,' 'high,' 'low,' 'falling,' 'rising,' 'creaky' and 'stopped' are adopted. The reasons for this choice can be summarized as follows.

(a) Vietnamese tones in all dialects display great similarities in pitch contour while differing in relative average pitch, and only one level tone occurs in each dialect system. This can be

TABLE 1.4 Frequency of Occurrence of Vietnamese Tones

Studies:	Emeneau 1951 *				My own 1978 **					
Tone	(a) n	%	(b) n	%	(c) NV n	%	(d) CV n	%	(e) SV n	%
Level /-/	418	24.4	368	36.8	2827	30.9	1841	28.6	3350	30.8
Falling /\`/	358	17.5	228	22.8	1998	21.8	1442	22.3	2179	20.0
Rising /ʼ/	548	26.8	180	18.0	2400	26.2	1836	28.5	3283	30.1
(/ʼs/)	(215)	(10.5)	(30)	(3.0)	(520)	(5.6)	(565)	(8.8)	(542)	(5.0)
Drop /./	333	16.3	144	14.4	956	10.5	604	9.4	964	8.9
(/.s/)	(128)	(6.2)	(38)	(3.8)	(381)	(4.1)	(253)	(3.9)	(423)	(3.9)
Curve /ʾ/	214	10.5	53	5.3	637	7.0	722	11.2	1106	10.2
Broken /~/	93	4.5	27	2.7	328	3.6				
Totals	2044	100.0	1000	100.0	9146	100.0	6445	100.0	10882	100.0

* Adapted from Emeneau 1951 (p 31 ff). (a) is based on the basic vocabulary (homonyms counted separately); (b) is based on "very colloquial conversations between a cook and various stallkeepers at market."

** My own study is based on informal conversations over a wide range of topics by groups of university students and staff. They involved (c) 7 NV, (d) 4 CV and (e) 9 SV speakers. The tones counted are phonological tones, with figures for the stopped variants in parentheses.

inferred from the phonetic notations in Table 1.3 (p 17) and will be shown in detail in the acoustic data presented in Ch. Two. This suggests that contour is a prominent feature in the Vietnamese tone system as a whole, and the unitary contour features are more appropriate than other systems. To take an example, the rising tone /´/ has the standard form [35] in NV and SV and [24] in CV, but their stopped variants can be [45] and [34], and there are other possible variants such as [25] or [14]. This shows that the feature [+ rising] underlies all these forms while the average pitch feature may vary between high and mid. A sequence-of-levels representation (as proposed in Leben 1978 for African tones or in Yip 1980 for Chinese tones, among others) would fail to capture this underlying feature in the Vietnamese system. The question remains open whether tones are of different types requiring different analyses.

(b) In this framework, 'contour' means 'not level' and 'falling' and 'rising' are self-explanatory. I have substituted 'concave' for 'convex' in Wang's system, for the reason that no convex tones occur in Vietnamese, and therefore 'convex' cannot serve to distinguish between the unidirectional and bidirectional tones in Vietnamese, where 'concave' can.

(c) The features 'high' and 'low' will establish three levels of average pitch for the phonological tone system. I have substituted 'low' for both 'mid' and 'central' in Wang's system because there is no need for more than three pitch levels at this level of representation in Vietnamese. The five levels of pitch targets needed for phonetic detail at onset, midpoint and endpoint are represented at the systematic phonetic level only. The three levels of average pitch will be supported by acoustic data given in Ch. Two (2.3.2).

(d) The features 'creaky' and 'stopped' are added to Wang's system for specific reasons. 'Creaky' is the voice quality feature corresponding to 'Laryng' at the phonetic level. It is the only feature that differentiates the NV falling tone and drop tone. Although it may be redundant, in a strictly distinctive feature system, with the NV broken tone and the CV curve tone, it remains an important characteristic of these tones. 'Stopped' is not strictly a tone feature, but rather a feature of the syllable on which the stopped tonesoccur. I use this feature to mark off the stopped variants, because, as noted earlier in 1.5.4.1, this feature underlies both their occurrence and their recognition as such by the hearer.

The full feature specification of standard NV, CV and SV tones are given in Table 1.5 (next page). Note that the features in parentheses are redundant and can be predicted by conventions or rules that I shall described in Ch. Five (section 5.1).

The use of the feature system described above is consistent with the basic assumption that tone is suprasegmental in nature and the syllable is the smallest tone-bearing unit, of which tone is a structural constituent as described in 1.5.3. This does not rule out the possibility that tonal phenomena can occur across syllable boundaries, in higher level units, and a theory of tonal phonology should make provisions for this too.

The foregoing points will be further discussed with supporting data at relevant parts of my study in accordance with the plan outlined in 1.3.

TABLE 1.5 Phonological Features of NV, CV and SV Tones

Dialect & Tone		(Con-tour)	(Con-cave)	Hi	Lo	Fa	Ri	Cr	(Stopped)*
NV Level /-/	˧	-	-	-	-	-	-	-	-
Falling /\`/	˧˩	+	-	-	+	+	-	-	-
Rising /'/	˧˥	+	-	+	-	-	+	-	-
St Ris /'s/	˥s	+	-	+	-	-	+	-	+
Drop /./	˧˩	+	-	-	+	+	-	+	-
St Drop /.s/	˩s	+	-	-	+	+	-	-	+
Curve /ʾ/	˨˩˨	+	+	-	+	+	+	-	-
Broken /~/	˧˩˥	+	+	+	-	+	+	+	-
CV Level /-/	˥	-	-	+	-	-	-	-	-
Falling /\`/	˥˧	+	-	-	-	+	-	-	-
Rising /'/	˧˦	+	-	-	-	-	+	-	-
St Ris /'s/	˧˦s	+	-	-	-	-	+	-	+
Drop /./	˧˩	+	-	-	+	+	-	-	-
St Drop /.s/	˩s	+	-	-	+	+	-	-	+
Curve /ʾ/	˨˩˨	+	+	-	-	+	+	+	-
SV Level /-/	˧	-	-	-	-	-	-	-	-
Falling /\`/	˧˩	+	-	-	+	+	-	-	-
Rising /'/	˧˥	+	-	+	-	-	+	-	-
St Ris /'s/	˥s	+	-	+	-	-	+	-	+
Drop /./	˨˩˨	+	+	-	+	+	+	-	-
St Drop /.s/	˩s	+	-	-	+	+	-	-	+
Curve /ʾ/	˨˩˦	+	+	-	-	+	+	-	-

* Features in parentheses are redundant and predictable by rules.

CHAPTER TWO

ACOUSTIC PHONETIC PROPERTIES OF VIETNAMESE TONES

2.0 *INTRODUCTION*

As noted in Ch. One, there have been a number of acoustic phonetic descriptions of Vietnamese tones. The earlier studies by Le Van Ly 1948, Andreev and Gordina 1957 and Nguyen Ham Duong 1962, based on kymograph or intonograph tracings, established the typical Fo contours of NV tones, regarded by the authors as representative of modern standard Vietnamese. Later studies by Han 1969, Han and Kim 1972, based on sound spectrograms, and by Earle 1975, using computerized pitch extractors, gave essentially similar results for NV tones. In these works, the authors mainly presented the Fo data and some data on duration, but no data on intensity. They mentioned laryngealization, or glottalization, as characterizing some tones but offered no detailed description.

The study of SV tones by Tran Huong Mai 1969 was partly based on pitch meter tracings, but no statistical data were provided. The author described SV tone contours in various stress and intonation patterns primarily from auditory study of recorded voices and visual inspection of the Fo tracings. To my knowledge, there has been no acoustic phonetic study of CV tones.

This chapter will deal with the physical phonetic aspects of Vietnamese tones based on acoustical data. The body of data gathered from a wide range of NV, CV and SV informants will provide a basis for determining the main characteristics of NV, CV and SV tones as produced by native speakers, and the range of regional and individual variations.

The results will be presented and discussed in two parts. In the first part, they are given in the form of arithmetical mean values of fundamental frequency, relative intensity, relative duration and characteristic degrees of laryngealization. Mean values are supposed to represent the average, typical parameters of the tones for the group of informants concerned, but they do not always convey linguistically significant information. As noted by Ladefoged (1975:255), "generally speaking, the phonetic characteristics of a sound cannot be determined by measuring the absolute values of the physical phenomena involved." For example, the SV and CV level tones have the following mean Fo values in Hertz at six measurement points, for the respective groups of informants:

(SV)	191	191	192	191	189	184	Hz
(CV)	184	182	185	187	188	186	Hz

Both tones can be described as having a level Fo contour, but the Fo values convey the false impression that the SV level tone has higher average pitch than the CV level tone. On the contrary, I shall demonstrate that, in terms of relative pitch within their own systems, the SV level tone is a mid tone and the CV level tone is a high tone.

That is why in the second part, the mean values will be converted into relative scales (percentage or decimal), by means of normalization procedures. This normalization, which in practice eliminates the non-linguistic characteristics of the tones, will yield results in the form of physical phonetic parameters in normalized values. These will be linguistically more relevant and more useful for comparing tones in different dialects and between speakers. For example, the above-mentioned level tone Fo values, converted into FD($\bar{F}$) - Fo Differential, in percent, relative to a mean $\bar{F}$ for each group of informants - will have the following values at six timepoints:

(SV)	4	4	4	4	3	1	FD($\bar{F}$) percent
(CV)	12	10	12	14	14	13	"

This gives a better idea of the relative pitch of the level tone in each system. 0 being the mean $\bar{F}$, the SV values indicate a pitch level slightly above the average, and CV values indicate a higher level still. Tones with negative FD($\bar{F}$) values would be automatically understood to belong to the low pitch range.

On the basis of the results in both mean values and normalized values, I shall determine the main characteristics of NV, CV and SV tones, their similarities and differences in various parameters, and discuss some implications for further linguistic analysis.

2.1 *PROCEDURES*

2.1.1 *Informants*

This study was based on the recorded voices of over forty native speakers of NV, CV and SV. For technical reasons, some of the recordings were not good enough for instrumental analysis, though they were useful for auditory study. Therefore, only thirty-four (11 NV, 12 CV and 11 SV informants) were used in the data analysis. They include 14 female and 20 male informants, coded F and M. Their birth places are indicated on the map in Figure 2.1 (next page), which would give a fair idea of how regional dialects were represented. The number in the codenames increases in the southward direction within each sex group and each dialect.

Of these informants, about two-thirds were university students aged between 18 and 30, the remaining being post-graduates and other staff of older age. Most of them were living away from where they had been born and had been exposed to other dialects to some extent. As a whole, their speech could then be characterized as educated and representative of the standard NV, CV and SV dialects, although traces of their local dialects were still present.

I took care to leave out those whose accent, judging by my ears, was not typical of their own dialects, either because of strong influences from others, or because of strong idiosyncratic or local features. I also decided to include a proportionately larger number of CV speakers in view of the fact that CV speech is thought to be more varied, less standardized and less well known than the other varieties.

2.1.2 Word Lists

I felt that there was no need to replicate the studies by Han 1969, which dealt almost exclusively with the single-vowel syllable types, and by Earle 1975, which dealt with a larger range of syllables involving all types of consonants and vowels. After

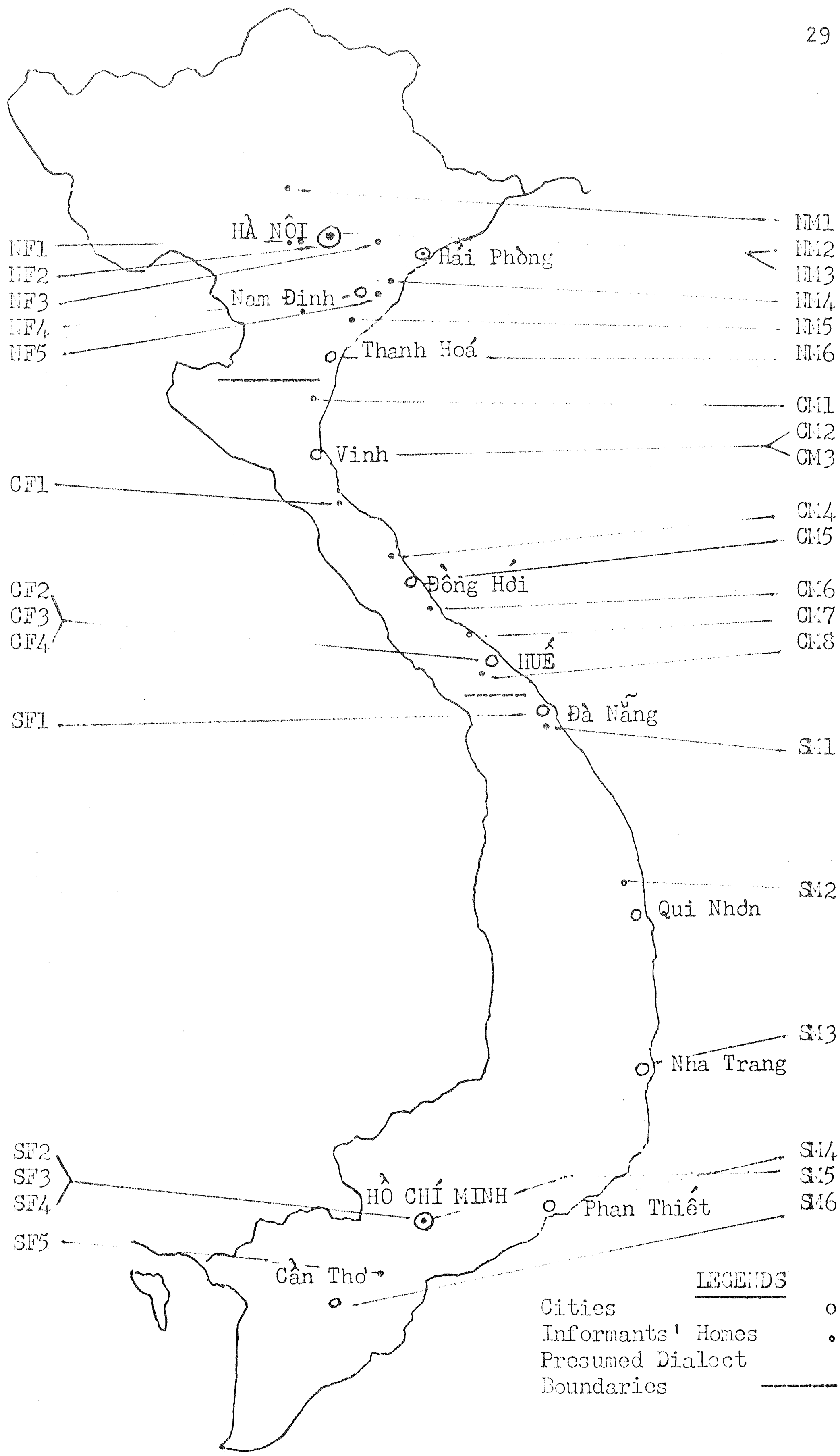


FIGURE 2.1 Distribution of Informants' Birth Places in Viet Nam

a pilot study of samples from informants available in Canberra had shown that the smaller samples gave essentially the same results without the greater variance occurring in larger samples, I decided that a cross-dialect study could include a smaller number of syllables with similar phonetic realizations in all three dialects.

The analysis in this chapter is based on two word-lists, given in Table 2.1 and Table 2.2 (next page). The first list contains five isolated syllables to be utterred in succession in the order given in the list, all with the vowel /a/ preceded by five different alveolar consonants (and for the stopped tones, followed by the velar /k/). All except a few occur as words or morphemes in Vietnamese, but their succession as in List A is not meaningful. The second list consists of the syllable /ta/ occurring with all non-stopped tones (and /tak/ for the stopped tones), repeated three times and preceded by a carrier frame which would make the utterances meaningful and more natural. The choice of these syllables, based on the fact that they represent some of the most neutral sounds that vary little across dialects, was intended to provide data typical of the average standard forms of tone in natural NV, CV and SV speech. Other word-lists were also given to informants to read; they will be described in relevant chapters.

2.1.3 Recordings

Voice recordings were made at different stages of my study and in somewhat different conditions, although the same procedures were followed. Of the thirty-four recordings retained for instrumental analysis, fifteen were made at the Phonetics Laboratory of the Department of Linguistics, The Faculties, of the Australian

TABLE 2.1 WORD LIST A

Tone	Syllables
Level /-/	/ tā tʰā dā nā sā /
Falling /\/	/ tà tʰà dà nà sà /
Rising /´/	/ tá tʰá dá ná sá /
Drop /./	/ tạ tʰạ dạ nạ sạ /
Curve /ʼ/	/ tả tʰả dả nả sả /
Broken /~/	/ tã tʰã dã nã sã /
Stopped Rising /´s/	/ tấk tʰấk dấk nấk sấk /
Stopped Drop /.s/	/ tạk tʰạk dạk nạk sạk /

TABLE 2.2 WORD LIST B

Tone	Frame	Syllables
Level /-/	/ ɛ̄ŋ ɛ̄m ' we brothers '	tā tā tā /
Falling /\/	/ bủəj cìəw 'declining evening'	tà tà tà /
Rising /´/	/ bán hāj 'sell two dozens'	tá tá tá /
Drop /./	/ nặŋ bā 'weighing three quintals'	tạ tạ tạ /
Curve /ʼ/	/ zẽ mō 'easy to describe'	tả tả tả /
Broken /~/	/ mɯ̄ə tɤ̀m 'raining cats and dogs'	tã tã tã /
Stopped Rising /´s/	/ bản ŋwīən 'original copy'	tấk tấk tấk /
Stopped Drop /.s/	/ zóŋ ɲɯ̄ 'resemble strikingly'	tạk tạk tạk /

<u>Note</u> The utterances, as enclosed between slanted strokes, were read in the order given above.

National University in Canberra, ten were made at a language laboratory of the Foreign Languages Institute in Hanoi, and the rest were made during my field trip to other parts of Vietnam. A UHER 4000 Reporter IC Recorder was used with reel-to-reel BASF tapes. Tape speed was set at 9.5 cm per second and volume control was set at Automatic. When the recordings were played back for study, I found no evidence that the amplitude had been clipped off.

Informants were first given instructions and the word lists to peruse then asked to read List A with some care (lower speed) and List B in a more natural way (faster speed). The idea was to have a corpus of syllables that represent the average standard tone forms in natural speech.

2.1.4 Instrumental Analysis

After a preliminary selection to eliminate poor recordings and achieve a fair balance of dialectal and individual representation for the study, the recorded materials were analyzed instrumentally. An F-J Fundamental Frequency Meter, an F-J Intensity Meter, a Sony 8-Channel Mixer and an Elema-Schönander Mingograf were used to analyze the output from the tape recorder and mingograms were obtained for a total of about 2000 syllables. The mingograms comprised Fo curves, intensity curves and duplex oscillograms of the oral waveforms. For the Fo analysis with the Fundamental Frequency Meter, the high-pass filter was set at 70 Hz and 120 Hz, and the low-pass filter at 200 Hz and 300 Hz, respectively for the male and female informants. For intensity recordings, volume control at the Mixer, set for each informant, ensured that peak intensities of most syllables varied between 5 and 10 dB, after the Mingograf controls had been calibrated with the Intensity Meter. I chose the full frequency range channel and linear display on the Intensity Meter. The calibration was the same for all informants so that results were comparable.

The mingograms were supplemented by about 300 narrow-band and wide-band spectrograms made from a Voiceprint Spectrograph over a number of syllables typical of each informant for checking and comparison in case of unclear tracings from the Mingograf. The spectrograms were made in linear bar display, with high speed and high shaping settings.

2.1.5 Measurements

The problem of analyzing the output from the Mingograf and the Spectrograph involves several decisions on what to measure and how to quantify and interpret the data. As previous studies mentioned in 2.0 have concurred in defining relative pitch, pitch contour and laryngealization as distinctive features for Vietnamese tones, I decided first to look at fundamental frequency as a major parameter of tone, then to make measurements of intensity and duration of a number of syllables with different tones, then to look at laryngealization.

2.1.5.1 Fo Measurements A preliminary visual examination of mingograms and spectrograms coupled with data from Han 1969 and Earle 1975 showed that the characteristic contour of any Vietnamese tones could be determined by measuring a certain number of points (three in those studies) on the Fo curves on the mingograms or a particular harmonic on a narrow-band spectrogram. For the purpose of this study, I decided that Fo curves on the mingograms were to be the primary objects of measurements, supplemented by spectrograms only in cases of unclear traces.

A selection of 8 syllables for each tone and each informant (five from List A and three from List B for each tone), was considered appropriate to represent the average standard forms. Therefore the number of syllables involved in characterizing Fo for each tone was 72 for NV and SV (nine informants for each dialect) and 96 for CV (twelve informants).

I also decided that a three-point measurement was not enough to represent some fine phonetic details characteristic of some tones (e.g. the lowest frequencies of the NV and CV curve tones usually occur at about two-thirds of the duration, while those of the SV curve tone and the NV broken tone usually occur at about one-third of the duration.) A ten-point measurement would yield graphic representations closely resembling real curves, but because of the amount of data to be measured by hand was too large for me to handle, I adopted a six-point measurement as a practicable compromise. The timepoints designated as P1, P2, P3, P4, P5 and P6 for non-stopped tones were defined respectively as occurring at onset, 10% (after-onset), 37% (midpoint 1), 63% (midpoint 2), 90% (before-endpoint) and endpoint. The timepoints defined for stopped tones, P1, P2, P3 and P4, occurred respectively at onset, 37% (midpoint 1), 63% (midpoint 2) and endpoint. The duration of stopped tones was found to be only between one-half and two-thirds of the average duration of other tones and the inclusion of measurements at 10% after onset and before endpoint would not yield any more significant data than had been obtained from four measurement points. Thus for each of the syllables measured, six Fo values were obtained for non-stopped tones and four Fo values for stopped tones.

The onset at P1 was taken to be the beginning of voicing in syllables except those with voiced consonant initials, in which case P1 coincided with the vowel onset as different from the consonant onset. Voiced consonant initials were excluded from the phonetic domain of tone for reasons discussed in Howie 1976, Sauvain 1977 and Rose 1981. The vowel onset at P1, in such cases, was determined from the characteristic dip of Fo during the obstruent /d/ (Lea 1973:61) or any appreciable discontinuity in either the Fo curve or the corresponding waveform and intensity curve. The endpoint was defined as the syllable end, whether it was the last voicing point of a vowel, a glide, a voiced consonant (which in Vietnamese can only be a nasal in this position) or the last measurable point of the laryngealized portion of the syllable. It appeared in the Fo curves under study (and also for syllables ending in glides and nasals not included in this chapter) that the characteristic Fo curve of each tone extends primarily over the vowel nucleus and sonorant ending of all syllables. When in doubt, both end point and onset were determined by comparing the Fo curve with the corresponding waveform , intensity curve or spectrogram. The intermediate points were measured at the defined timepoints when the curve was smooth, but when it was irregular, the nearest turning point, peak or low point was measured instead. When there were Fo perturbations due to various reasons (usually breathy voice, creaky voice or consonantal release), the nearest stable point was measured, or if the perturbation was of considerable duration, a point was taken by interpolation between two stable points. Examples of those points can be found in Figures 2.2, 2.3, 2.4 and 2.5 (pp 40 to 43).

2.1.5.2 Intensity and Duration Measurements Most studies of Vietnamese tones noted in 2.0 barely mentioned duration as characteristic of some tones (the stopped tones, the NV creaky drop tone) and to my knowledge no data on intensity are available. My data on intensity and duration were extracted from the same syllables, /ta/ for non-stopped tones and /tak/ for stopped tones, so as to eliminate variations that might have been associated with possible differences in the intrinsic intensity and duration of various types of vowels and consonants. Intensity was measured from four syllables for each tone and each informant, at four time-points, designated by I1, I2, I3 and I4, corresponding to after-onset, midpoint 1, midpoint 2 and before-endpoint of the Fo measurements. There was no need to measure intensity at onset and endpoint because in the type of syllable chosen, beginning with a voiceless /t/ and ending in /a/ or /k/, it was found to be zero or near zero at those points.

Duration was measured in centiseconds and the determination of onset and endpoint was made in the same way as for Fo and intensity.

2.1.5.3 Statistical Treatment The Fo, intensity and duration values obtained from measurements described in the previous subsections were treated statistically by taking the arithmetic means of all tokens of a tone sample for each informant and for informants of the same dialect group, and standard deviations were calculated.

The standard deviation is a measure of dispersion of the statistical data, while the arithmetic mean, the conventional 'average', is a measure of central tendency (MacDonald 1977).

The statistical formula for calculating the mean is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

where $\bar{x}$ is the mean,

$x_i \ldots\ldots x_n$ are tokens in the set of data, and

n is the number of tokens in the sample.

The standard deviation is calculated by taking the square root of the sample variance S^2, which is defined statistically as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

The value n-1 is used instead of n in the variance formula because the variance is taken as an estimate from the sample and not as the exact variance of the population. The standard deviation is normally represented by the symbol S in statistical formulae, but it will be abbreviated as SD in my Tables.

2.1.5.4 Laryngealization All previous studies of NV tones mentioned laryngealization as a feature of the drop tone and the broken tone, under a variety of terms such as laryngealized, glottalized, constricted, creaky, glottal stricture, etc. Breathy voice was also mentioned as an optional feature of the curve tone. In one of the later acoustic studies, Earle 1975 used the term 'laryngeal effect' to cover creaky voice, breathy voice and glottal closure occurring in NV tones, but gave no quantified data.

As there has been no tested method of quantifying laryngealization, as I defined in Ch. One (1.2), I decided to combine auditory study with observation of perturbations on spectrograms, Fo curves and waveforms (such as irregular striae on wide-band and blurred harmonics on narrow-band spectrograms, or sudden drops or jumps on Fo curves or any irregular forms of the waves, i.e. anything that suggests the presence of irregular voicing.) The presence of creaky voice and breathy voice in any portion of a tone was noted when it occurred fairly regularly for a particular tone over the sample of syllables by the whole group of informants for NV, CV or SV (dialectal characteristic) or by a particular informant (individual characteristic). Duration of the laryngealized parts was measured roughly on spectrograms of the same syllables /ta/ and /tak/ used in the intensity and duration measurements, so that possible variations associated with different types of consonants or vowels were not involved.

Illustrative mingograms and spectrograms of syllables with NV tones by Informant NF3 are given in Figures 2.2 to 2.5 (following pages) in which some measurement points and laryngealized parts are indicated.

2.2 *RESULTS*

The results are given in tabular forms in the following pages, in the order Fo, intensity, duration and laryngealization.

Tables 2.3 to 2.5 present the mean Fo and standard deviations of NV, CV and SV tones respectively at six timepoints (four timepoints for stopped tones); the values are in Hertz and rounded to the nearest unit. Details of mean Fo values of each tone for each informant can be found in the Appendix to this chapter, together with more illustrative mingograms and spectrograms for various tones in the three dialects.

Tables 2.6 to 2.8 present the mean intensity values and standard deviations of NV, CV and SV tones respectively at four time points; the values are in dB and given with one decimal place.

Table 2.9 presents the mean duration values in centiseconds and standard deviations of NV, CV and SV tones together. Figures were rounded to the nearest centiseconds ; a maximum and a mimimum value from the 32 tokens in each tone sample are also included in the table for comparison.

The results of investigation into the occurrence of breathy voice and creaky voice are summarized in Table 2.10 in which four degrees of laryngealization are posited for categorization at the auditory level, namely

0 regular voicing, no noticeable change in voice quality;

1 breathy voice or slight laryngealization, usually associated with low tones by speakers having low Fo ranges, as free variant of 0 or alternative realization of 2;

2 creaky voice or heavy laryngealization;

3 glottal closure, usually an alternative realization of 2.

Following the above tables, graphic representations of the results are given in Figures 2.6 to 2.9.

Figure 2.6 plots the mean values of NV, CV and SV tones in their respective systems, on the same Fo scale, as functions of

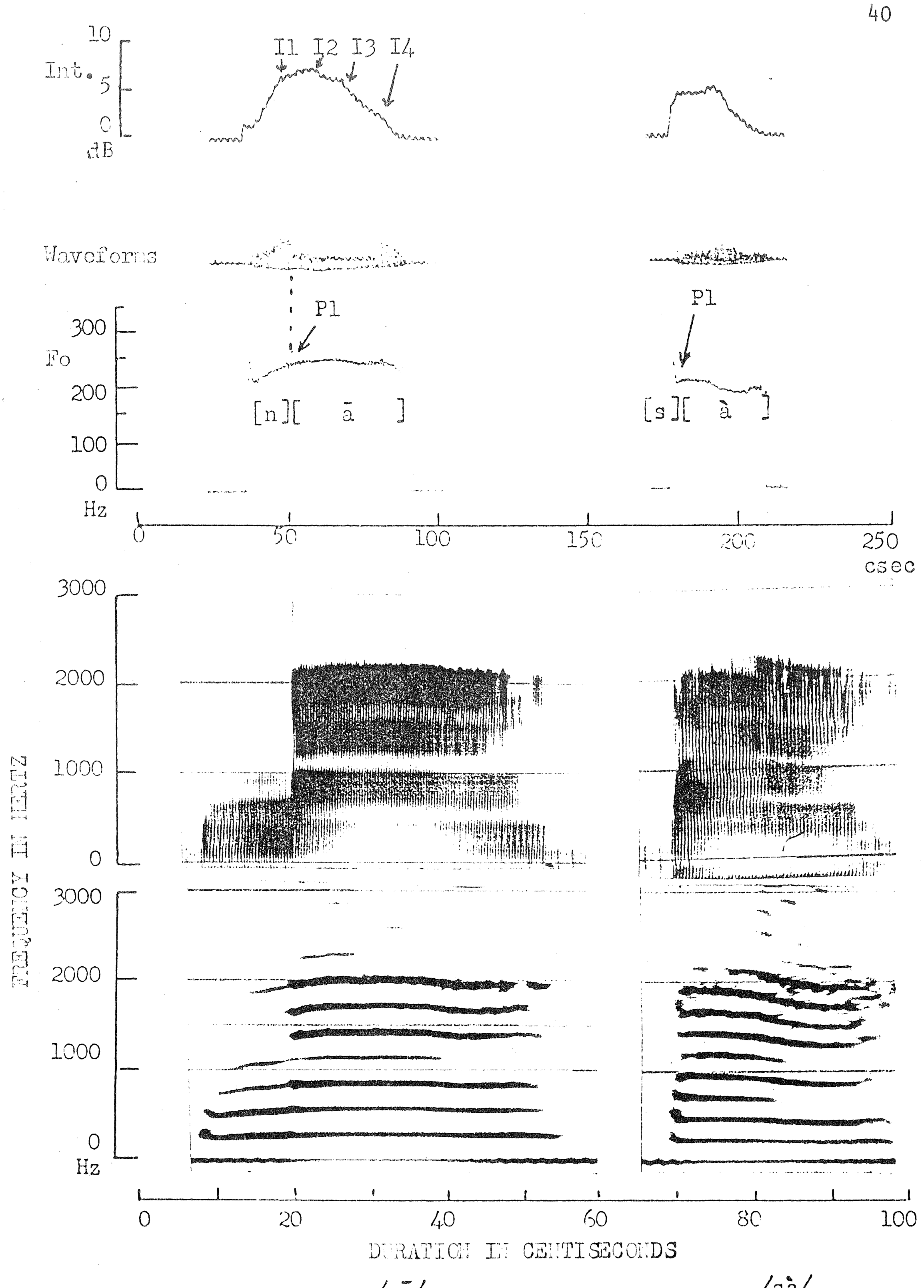


FIGURE 2.2 Typical Intensity Curves, Waveforms, Fo Curves, Wideband and Narrowband Spectrograms of Syllables with NV Level and Falling Tones by Informant MF3. (I and P indicate measurement points for intensity and Fo respectively.)

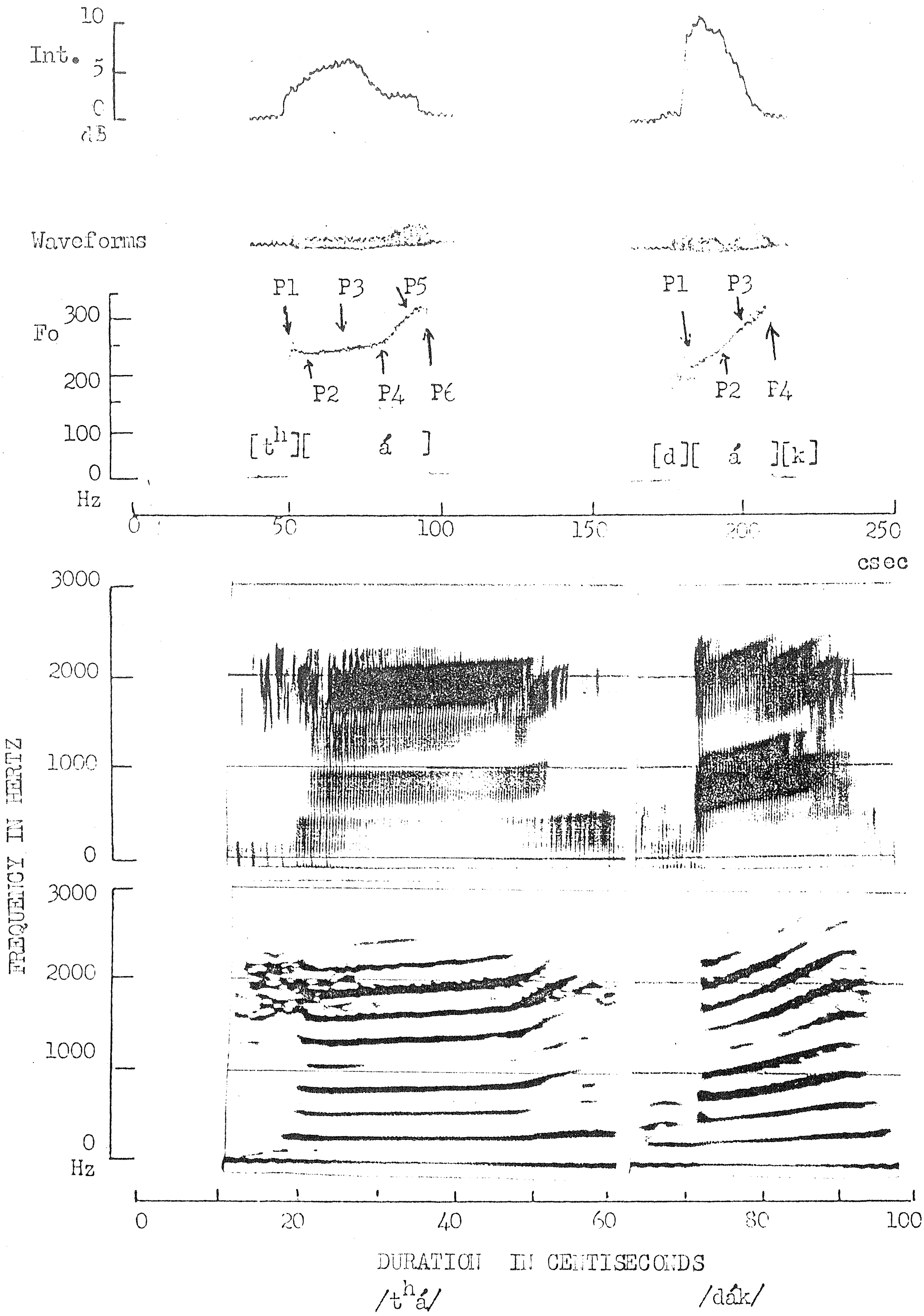


FIGURE 2.3 Typical Intensity Curves, Waveforms, Fo Curves, Wideband and Narrowband Spectrograms of Syllables with NV Rising and Stopped Rising Tones by Informant NF3. (Ps indicate measurement points for Fo.)

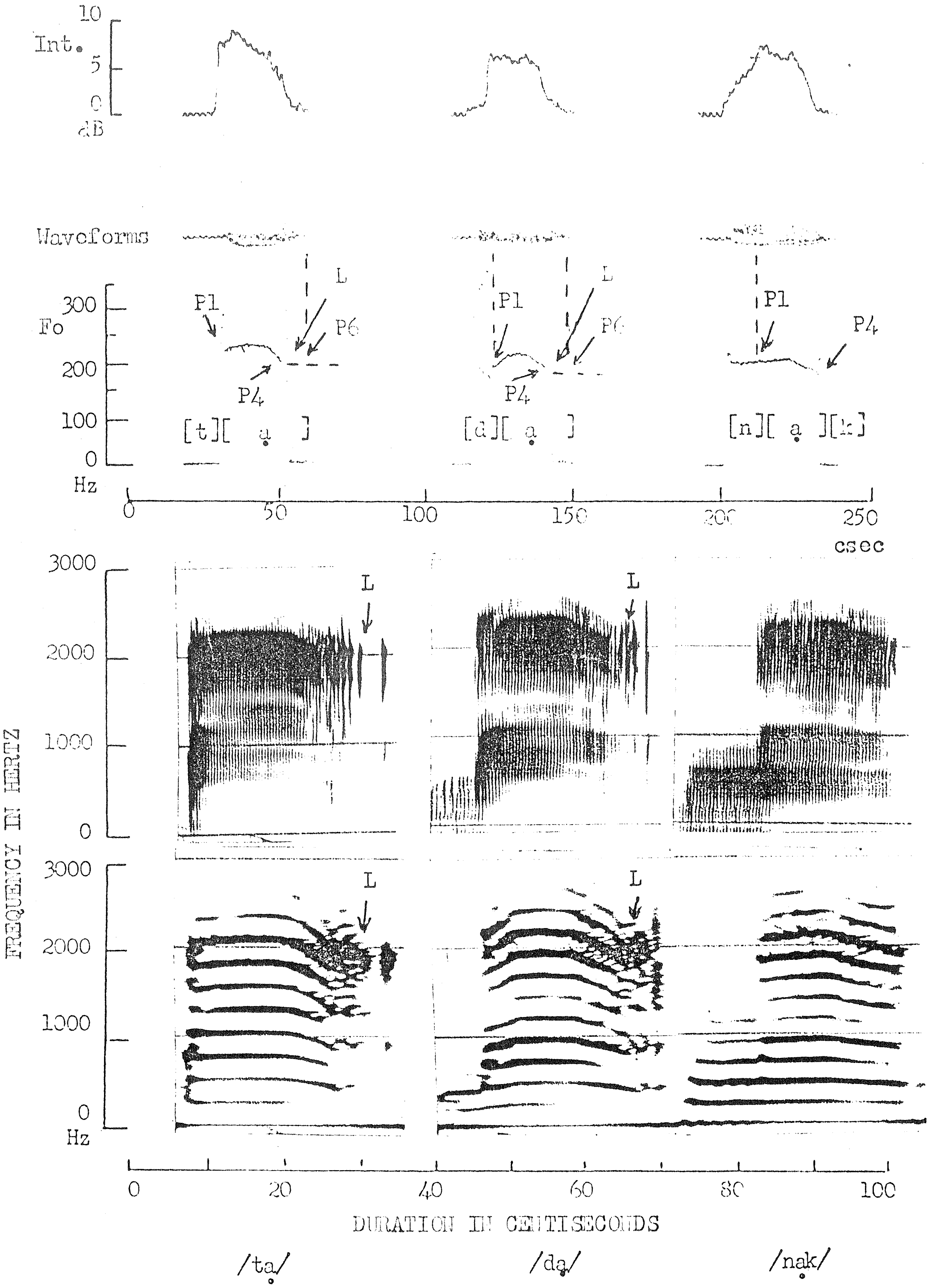


FIGURE 2.4 Typical Intensity Curves, Waveforms, Fo Curves, Wideband and Narrowband Spectrograms of Syllables with NV Drop and Stopped Drop Tones by Informant NF3. (Ps indicate measurement points for Fo, Ls point to laryngealization.)

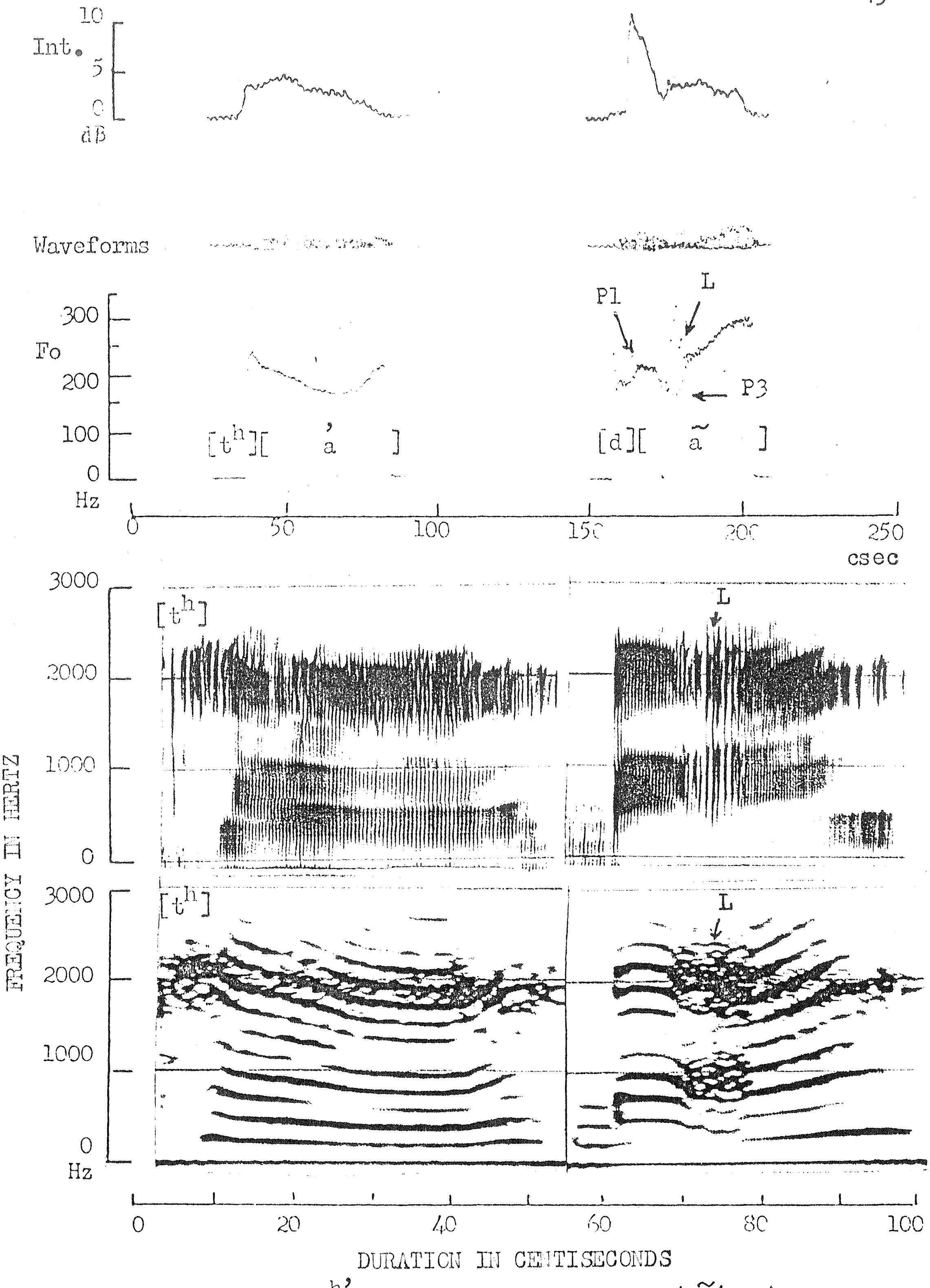


FIGURE 2.5 Typical Intensity Curves, Waveforms, Fo Curves, Wideband and Narrowband Spectrograms of Syllables with NV Curve and Broken Tones by Informant NF3. (Ps indicate measurement points for Fo, Ls point to laryngealization.)

TABLE 2.3 Mean Fo in Hertz and Standard Deviations of HV Tones at Six Timepoints (Four Timepoints for Stopped Tones)*

Tone	n	P1	P2	P3	P4	P5	P6
Level /-/	72	212	212	212	210	207	202
		55	56	55	55	54	53**
Falling /`/	72	178	175	171	163	159	155
		45	46	45	43	44	43
Rising /´/	72	198	194	197	217	246	257
		52	51	51	51	54	56
St. Ris. /´s/	72	221	232	250	268		
		61	59	57	60		
Drop /./	72	189	186	180	165	162	163
		47	47	45	43	44	47
St. Drop /.s/	72	182	174	167	158		
		45	43	41	43		
Curve /ˀ/	72	176	170	157	140	155	166
		45	46	42	37	38	38
Broken /~/	72	202	196	169	211	244	245
		55	54	47	55	55	58

* Data from nine HV informants: HF1, HF2, HF3, HF4, HM1, HM2, HM3, HM4 and HM6.

** For each tone, mean Fo values on first line, SD on second line.

TABLE 2.4 Mean Fo in Hertz and Standard Deviations of CV Tones at Six Timepoints (Four Timepoints for Stopped Tones)*

Tone	n	P1	P2	P3	P4	P5	P6
Level /-/	96	184	182	185	187	188	186
		53	50	50	50	52	52**
Falling /ˎ/	96	174	171	167	162	159	156
		49	48	47	46	45	44
Rising /ˊ/	96	158	153	150	162	177	181
		43	39	37	37	47	52
St. Ris. /ˊs/	96	163	159	166	179		
		42	39	39	46		
Drop /./	96	163	158	153	148	148	149
		42	40	38	39	39	40
St. Drop /.s/	96	169	161	155	150		
		43	39	39	39		
Curve /ˀ/	96	166	163	156	148	153	156
		45	43	41	42	46	48

* Data from twelve CV informants: CF1, CF2, CF3, CF4, CM1, CM2, CM3, CM4, CM5, CM6, CM7 and CM8.

** For each tone, mean Fo values on first line, SD on second line.

TABLE 2.5 Mean Fo in Hertz and Standard Deviations of SV Tones at Six Timepoints (Four Timepoints for Stopped Tones)*

Tone	n	P1	P2	P3	P4	P5	P6
Level /-/	72	191	191	192	191	189	185
		61	60	61	60	61	61**
Falling /\/	72	166	161	155	151	149	149
		52	52	51	50	51	54
Rising /'/	72	194	192	201	224	250	255
		63	60	61	68	80	84
St. Ris. /'s/	72	201	208	228	256		
		67	66	71	86		
Drop /./	72	166	157	148	150	162	166
		57	52	49	51	55	56
St. Drop /.s/	72	170	159	155	160		
		56	52	49	52		
Curve /?/	72	173	162	149	175	219	224
		61	56	48	54	66	71

* Data from nine SV informants: SF1, SF2, SF3, SF5, SM1, SM2, SM4, SM5 and SM6.

** For each tone, mean Fo values on first line, SD on second line.

TABLE 2.6 Mean Intensity in dB and Standard Deviations of NV Tones on Same Syllables at Four Timepoints*

Tone	n	I1	SD	I2	SD	I3	SD	I4	SD
Level /-/	32	7.1		6.3		5.4		2.0	
			2.1		1.1		1.6		1.8**
Falling /\/	32	5.1		4.5		3.5		1.1	
			2.1		1.5		1.7		0.2
Rising /'/	32	5.5		4.6		5.4		1.6	
			1.5		1.3		2.5		1.1
St. Ris. /'s/	32	5.6		5.0		4.5		1.5	
			2.0		1.6		2.1		1.3
Drop /./	32	4.5		4.1		2.3		0.8	
			1.3		1.5		1.0		0.2
St. Drop /.s/	32	5.0		4.4		3.6		1.0	
			2.1		1.8		1.6		0.05
Curve /'/	32	4.4		3.2		2.5		1.4	
			1.5		1.2		1.2		0.5
Broken /~/	32	5.2		2.4		4.8		1.4	
			1.9		0.9		2.3		0.5

* Data from 8 NV informants: NF1, NF2, NF5, NM1, NM3, NM4, NM5 and NM6. The syllables were /ta/ and /tak/ for sonorant-ending and stopped tones respectively.

** For each tone, mean intensity values on first line, SD on second line.

TABLE 2.7 Mean Intensity in dB and Standard Deviations of CV Tones on Same Syllables at Four Timepoints*

Tone	n	I1	SD	I2	SD	I3	SD	I4	SD
Level /-/	32	8.0		7.2		5.5		2.7	
			2.5		1.1		1.8		1.3**
Falling /ˎ/	32	8.4		6.5		4.1		1.9	
			2.7		1.8		1.6		0.6
Rising /'/	32	5.9		5.0		5.1		2.3	
			2.8		1.8		1.8		1.2
St. Ris. /'s/	32	5.3		4.9		4.6		1.6	
			2.6		2.1		1.8		0.5
Drop /./	32	6.5		4.7		3.1		1.7	
			3.1		1.5		1.9		0.9
St. Drop /.s/	32	6.6		5.3		3.8		1.4	
			3.2		1.5		1.3		0.8
Curve /ˀ/	32	7.3		5.8		3.2		1.4	
			2.7		2.2		1.6		0.7

* Data from 8 CV informants: CF1, CF2, CF4, CM1, CM4, CM5, CM7 and CM8. The syllables were /ta/ and /tak/ for sonorant-ending and stopped tones respectively.

** For each tone, mean intensity values on first line, SD on second line.

TABLE 2.8 Mean Intensity in dB and Standard Deviations of SV Tones on Same Syllables at Four Timepoints*

Tone	n	I1	SD	I2	SD	I3	SD	I4	SD
Level /-/	32	4.9		5.8		4.1		2.3	
			2.0		1.4		1.3		1.7**
Falling /\/	32	3.5		3.8		2.4		1.2	
			1.3		0.9		0.8		0.6
Rising /'/	32	4.4		4.9		4.2		2.5	
			1.8		1.3		2.2		2.3
St. Ris. /'s/	32	4.2		5.8		5.0		2.4	
			2.3		2.0		1.5		1.7
Drop /./	32	3.2		2.9		2.0		1.2	
			1.3		1.1		0.8		0.4
St. Drop /.s/	32	3.9		3.7		2.7		1.2	
			1.2		1.2		0.3		0.4
Curve /?/	32	3.6		2.5		3.0		2.0	
			1.5		0.9		1.0		1.7

* Data from 8 SV informants: SF1, SF4, SF5, SM1, SM3, SM4, SM5 and SM6. The syllables were /ta/ and tak/ for sonorant-ending and stopped tones respectively.

** For each tone, mean intensity values on first line, SD on second line.

TABLE 2.9 Mean Duration in Centiseconds and Standard Deviations of NV, CV and SV Tones on Same Syllables*

Dialect & Tone	n	$\bar{D}$	SD	Dmax	Dmin**
NV Level /-/	32	25	6	36	12
Falling /`/	32	25	6	38	12
Rising /'/	32	25	5	38	14
St. Rising /'s/	32	15	3	22	10
Drop /./	32	20	4	30	14
St. Drop /.s/	32	15	3	22	8
Curve /’/	32	26	6	40	16
Broken /~/	32	25	4	32	14
CV Level /-/	32	26	5	40	16
Falling /`/	32	28	5	40	16
Rising /'/	32	26	5	40	18
St. Rising /'s/	32	16	3	26	10
Drop /./	32	28	7	50	18
St. Drop /.s/	32	17	3	26	10
Curve /’/	32	24	5	40	15
SV Level /-/	32	30	8	50	16
Falling /`/	32	29	6	44	18
Rising /'/	32	28	6	40	15
St. Rising /'s/	32	18	4	28	12
Drop /./	32	28	5	40	18
St. Drop /.s/	32	17	4	24	10
Curve /’/	32	30	6	42	20

* Data from same syllables and same informants as for intensity data in Tables 2.6 to 2.8 (previous pages).

** Dmax and Dmin are the longest and shortest tokens found in each tone sample.

TABLE 2.10 Laryngealization in MV, CV and SV Tones *

Dialect & Tone	Degree (a)	Duration (b)	Timing (c)
MV Level /-/, Rising /ʹ/, St. Rising /ʹs/ & St. Drop /.s/	0		
Falling /ˎ/	0 (1)	(i)	(E)
Drop /./	2 (3)	4.5	E
Curve /ʾ/	0 (1)	(i)	(M,E)
Broken /~/	2 (3)	3.7	M
CV Level /-/, Falling /ˎ/, St. Rising /ʹs/ & St. Drop /.s/	0		
Rising /ʹ/	0 (1)	(i)	(M)
Drop /./	0 (1,2)	(i)	(E)
Curve /ʾ/	2 (1)	4.6	E
SV Level /-/, Rising /ʹ/, St. Rising /ʹs/ & St. Drop /.s/	0		
Falling /ˎ/	0 (1)	(i)	(E)
Drop /./	0 (1,2)	(i)	(E)
Curve /ʾ/	0 (1)	(i)	(M)

* Based on auditory and acoustic studies on same syllables and same informants as for intensity data in Tables 2.6 to 2.8 (previous pages).

(a) 0:regular voicing; 1:breathy voice; 2:creaky voice; 3:glottal closure. ( )indicates alternative occurrences with some speakers and in some contexts only.

(b) Number indicates mean duration of laryngealized part in centi-seconds; (i) indicates irregular durations.

(c) Laryngealization can usually occur at the middle M or end E of the syllable.

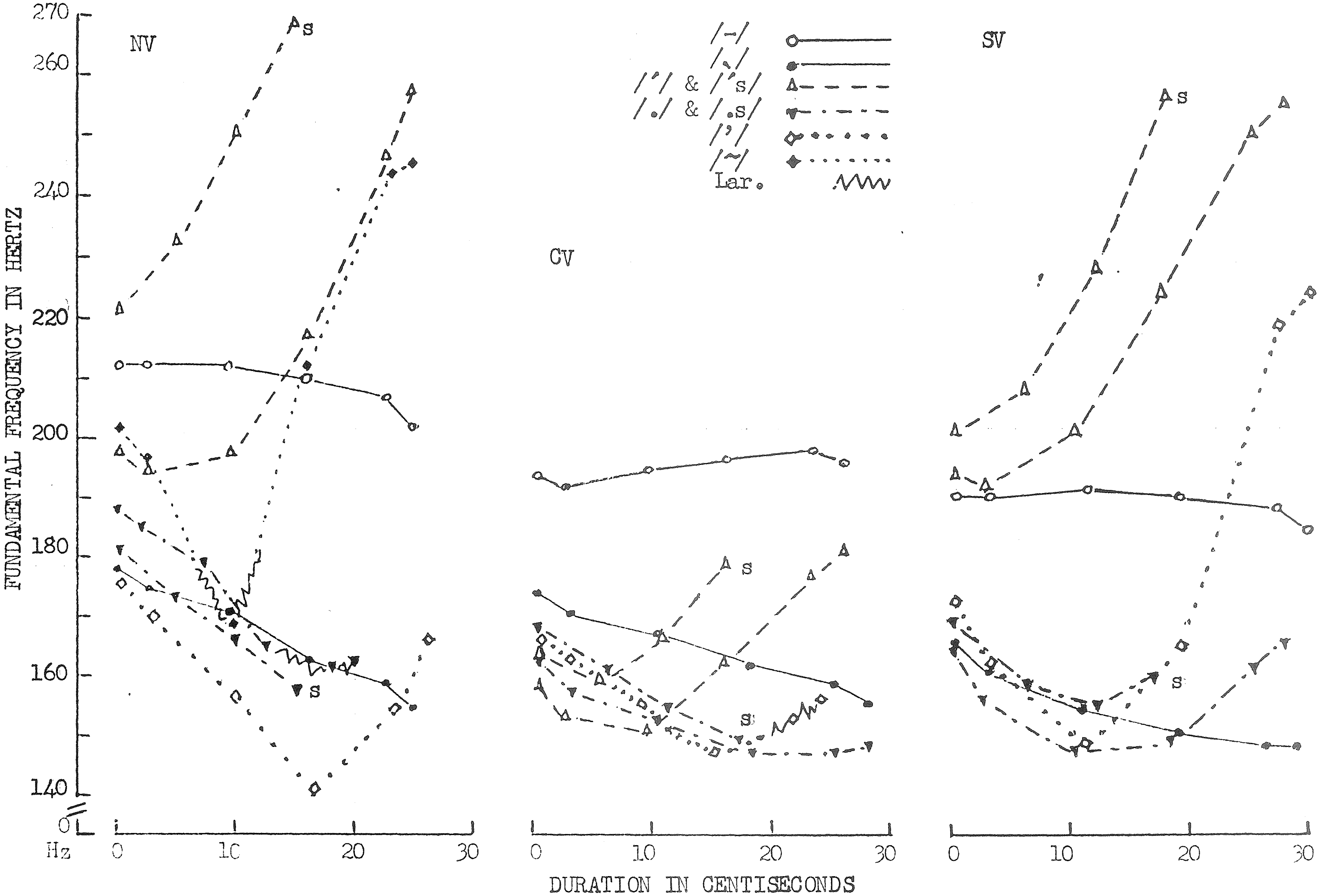


FIGURE 2.6 Mean Fo of NV, CV and SV Tones Plotted Against Mean Duration (Data from 9 NV, 12 CV and 9 SV informants).

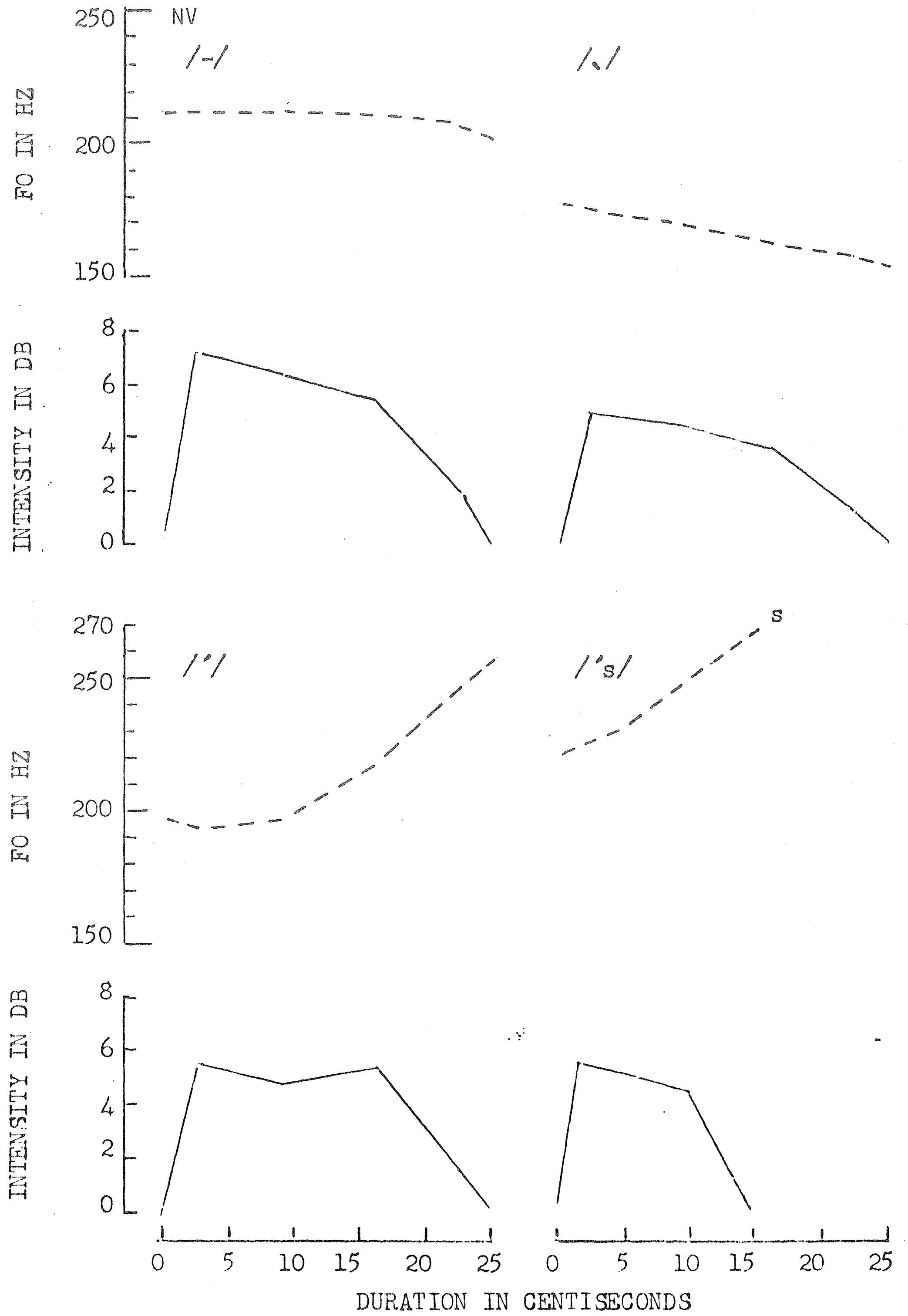


FIGURE 2.7(a) Mean Fo in Hertz and Mean Intensity in dB of NV Tones Plotted Against Mean Duration (Data from Tables 2.3, 2.6 and 2.9)

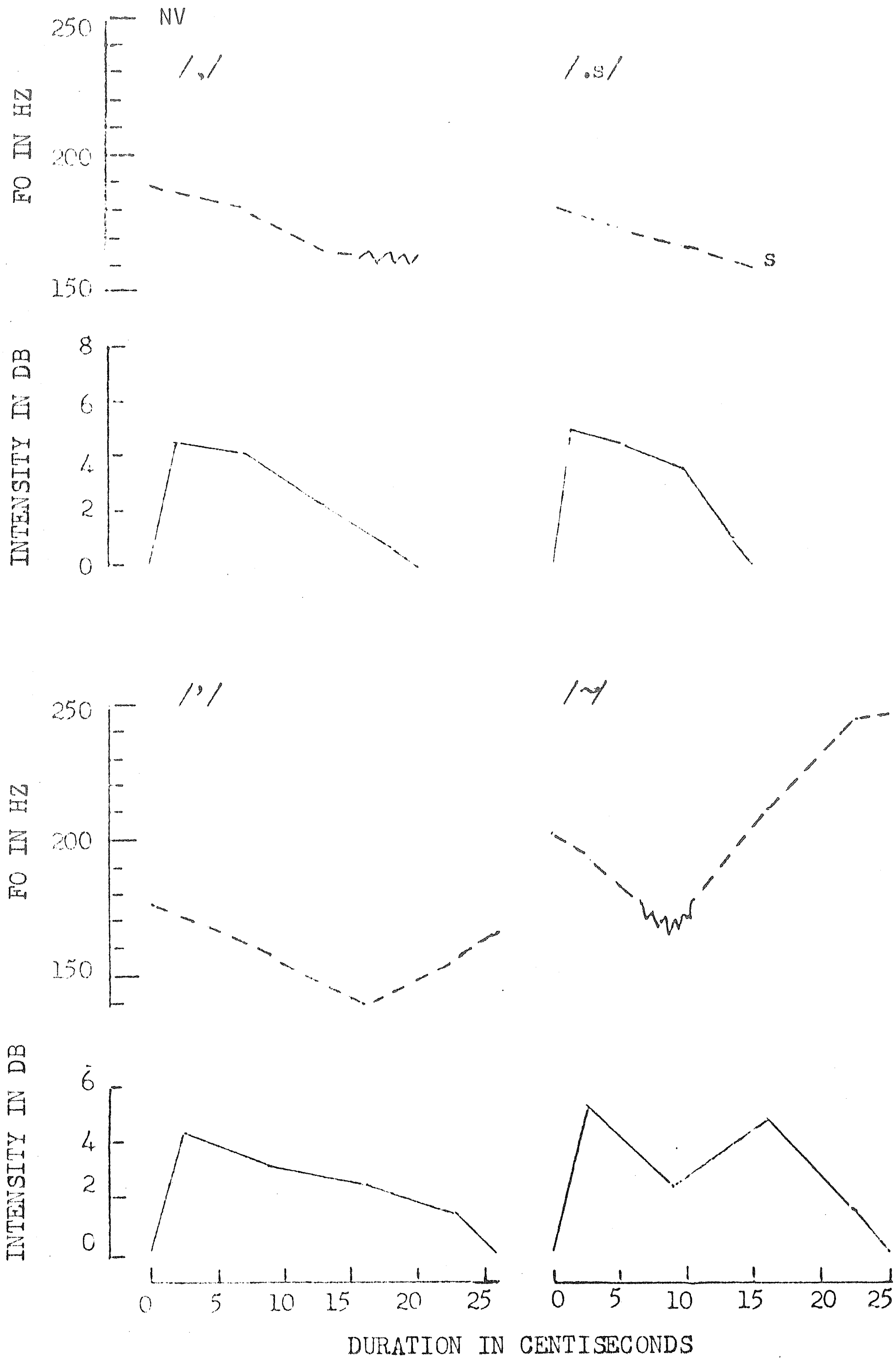


FIGURE 2.7(b) Mean Fo in Hertz and Mean Intensity in dB of NV Tones Plotted Against Mean Duration (Data from Tables 2.3, 2.6 and 2.9)

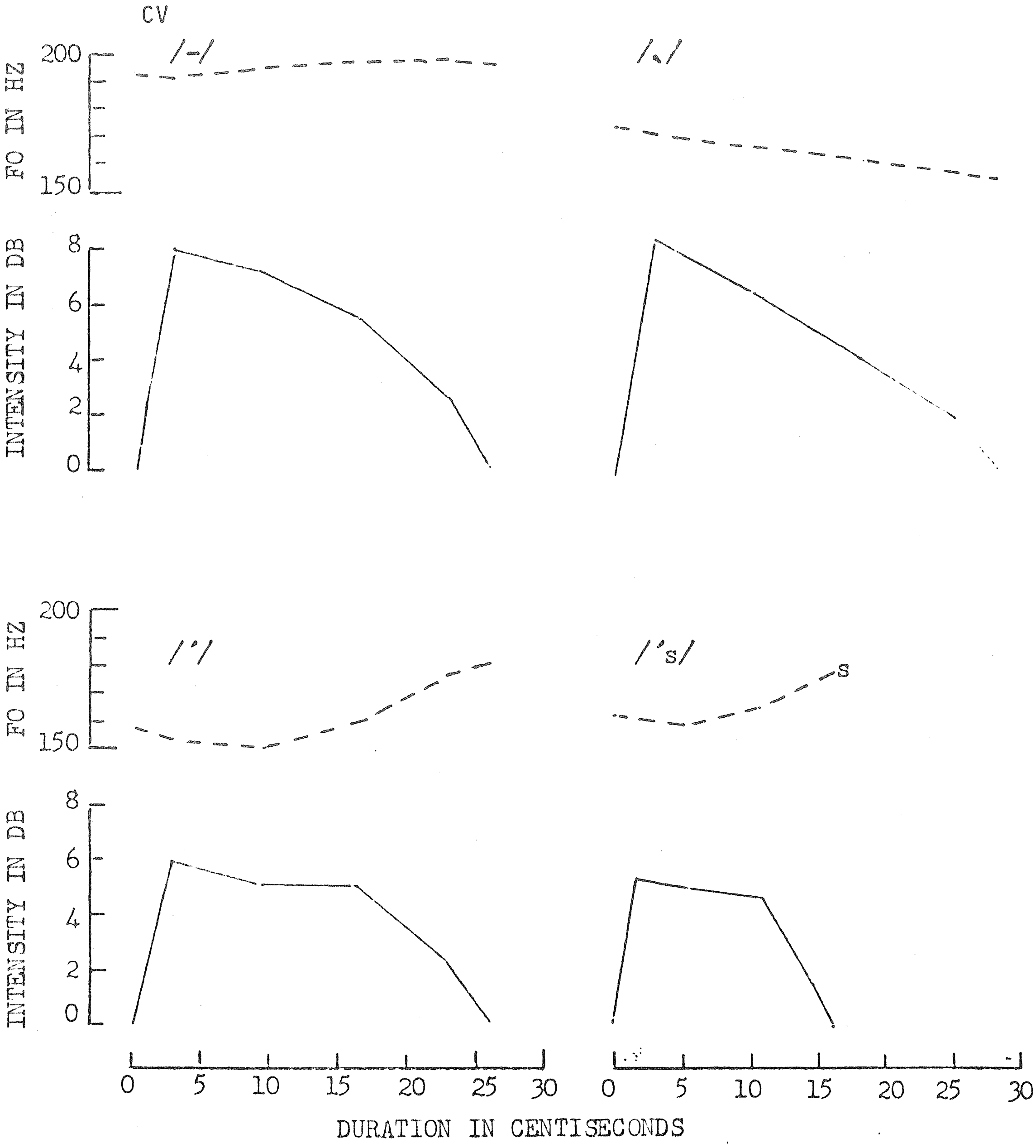


FIGURE 2.8(a) Mean Fo in Hertz and Mean Intensity in dB of CV Tones Plotted Against Mean Duration (Data from Tables 2.4, 2.7 and 2.9)

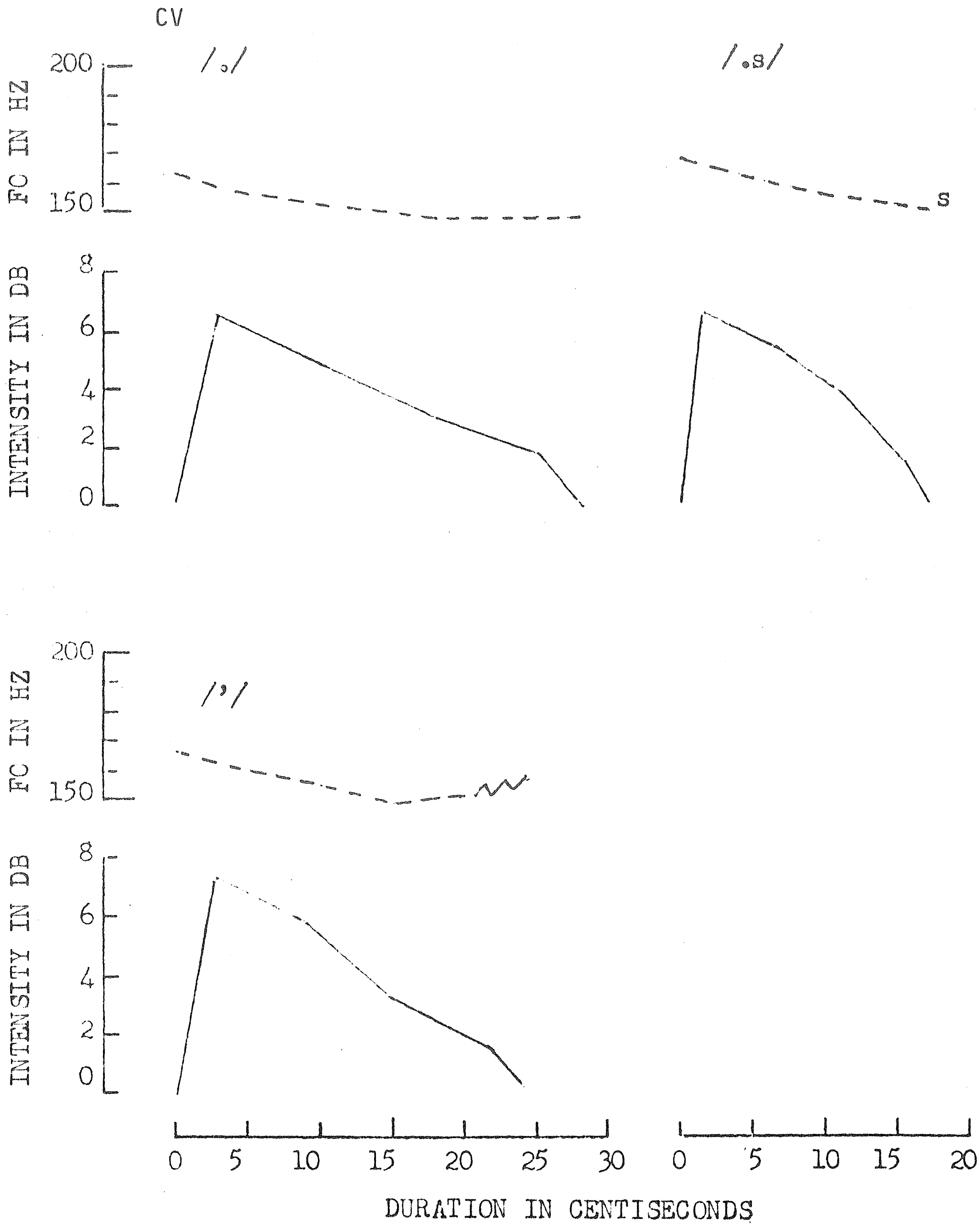
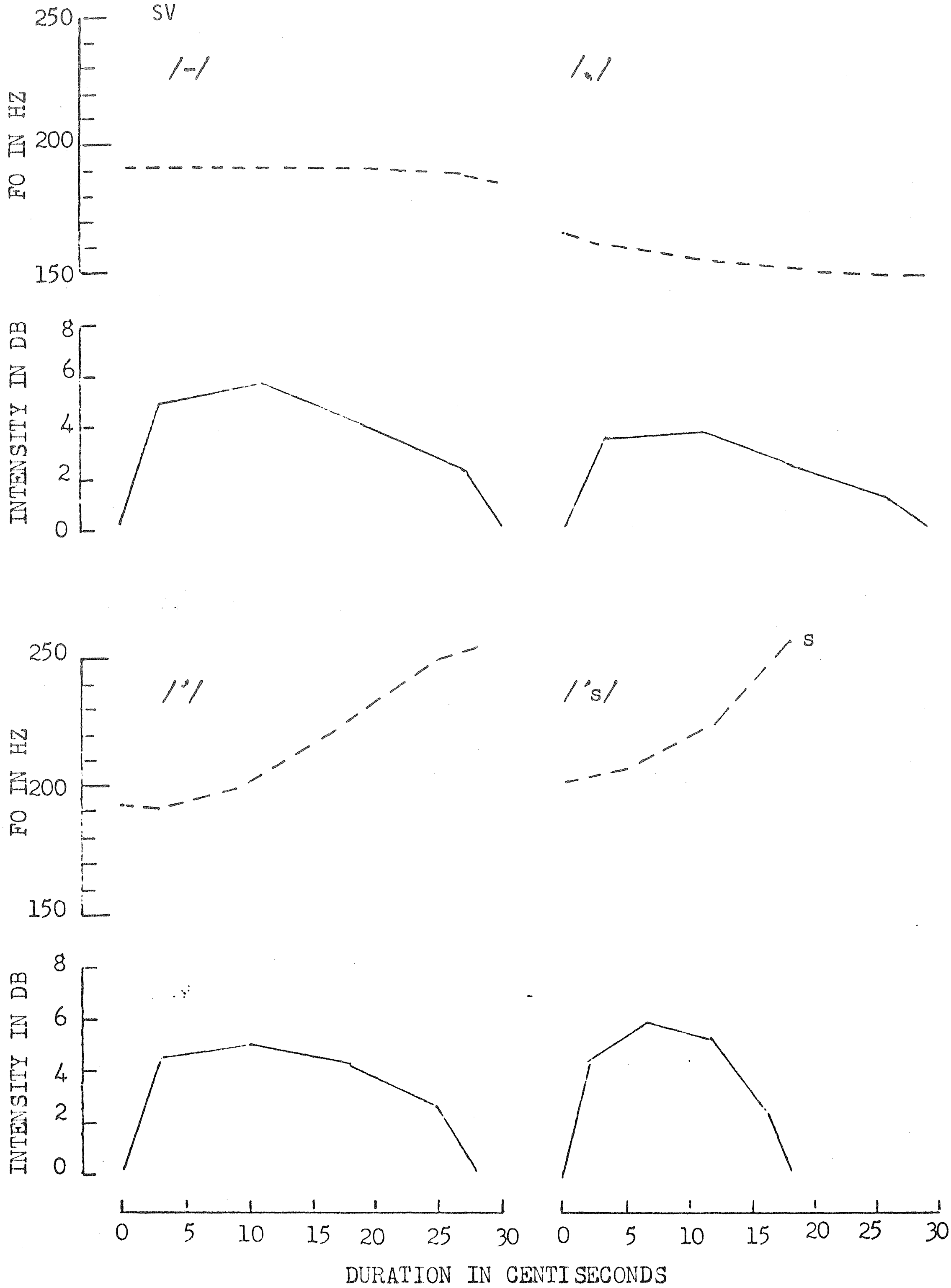


FIGURE 2.8(b) Mean Fo in Hertz and Mean Intensity in dB of CV Tones Plotted Against Mean Duration (Data from Tables 2.4, 2.7 and 2.9)



FIGYRE 2.9(a) Mean Fo in Hertz and Mean Intensity in dB of SV Tones Plotted Against Mean Duration (Data from Tables 2.5, 2.8 and 2.9)

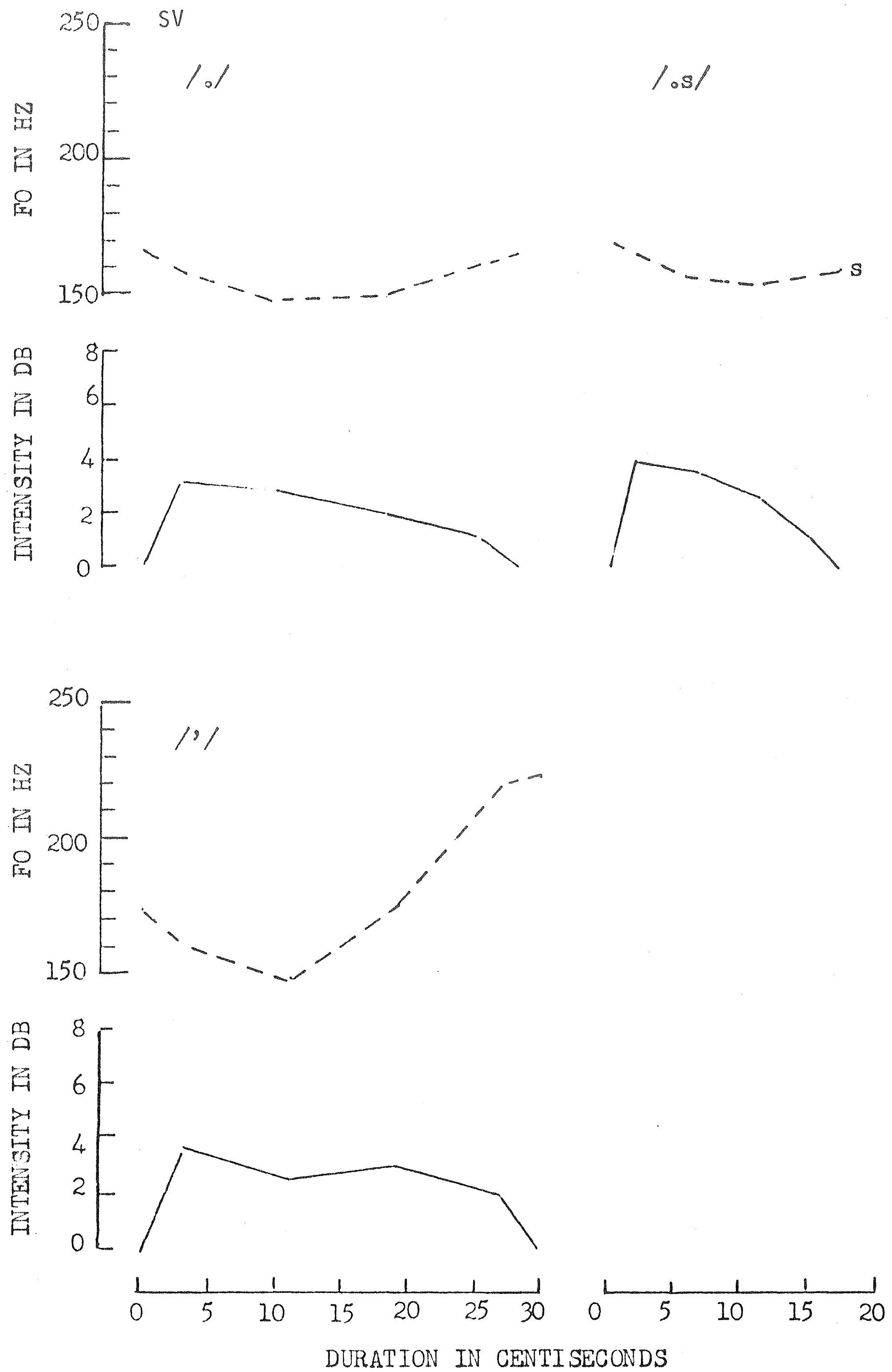


FIGURE 2.9(b) Mean Fo in Hertz and Mean Intensity in dB of SV Tones Plotted Against Mean Duration (Data from Tables 2.5, 2.8 and 2.9)

mean duration.

Figures 2.7 (a,b), 2.8 (a,b) and 2.9 (a,b) present synchronized graphs of the mean Fo values and mean intensity values of each tone of the three dialects, as functions of mean duration.

2.3 DISCUSSION OF RESULTS

In the following discussion, I shall make general observations on the acoustic characteristics of Vietnamese tones in the three major dialects as can be inferred from the data. Then I shall discuss particular points concerning the parameters Fo, relative intensity, duration and laryngealization.

2.3.1 Acoustic Characteristics of NV, CV and SV Tones

The following description of the three dialect systems can be made from inspection of the quantified data presented in the tables and figures in the preceding section. Before formal normalization procedures are introduced in the next section, let us assume that the labels 'level,' 'falling,' 'rising' and 'concave' in Fo contour can be impressionistically determined from the presence of appreciable changes in Fo between onset and endpoint; similarly, the labels 'sustained,' 'falling,' and 'concave' in intensity contour will be used depending on whether there are any appreciable differences between the peak intensity (usually between 1/10 and 1/3 of duration) and the third measurement point (at 2/3 of durration) when intensity generally starts to decay in all tones.

(a) <u>NV Tones</u>

<u>NV Level Tone /-/</u> Average Fo level: mid; level Fo contour

slightly falling toward the end; falling intensity contour; long duration; no laryngealization.

NV Falling Tone /\/ Average Fo level: low; falling Fo contour; falling intensity contour; long duration; no characteristic laryngealization, but sometimes breathy voice with speakers having low Fo ranges.

NV Rising Tone /ˊ/ Average Fo level: high; rising Fo contour, fairly level for about 1/3 of duration then starting to rise and reaching top Fo range toward the end; sustained intensity contour, fairly level between 1/10 and 2/3 of duration, then starting to decay well before peak Fo is reached; long duration; no characteristic laryngealization.

NV Stopped Rising Tone /ś/ Average Fo level : higher than the non-stopped rising tone; Fo onset starts higher and rises sharper within a shorter duration of about 3/5 of the non-stopped tone; sustained intensity contour; no laryngealization.

NV Drop Tone /./ Average Fo level: low; falling Fo contour starting and ending a little higher than the falling tone; falling intensity contour; slightly shorter duration (4/5 of falling tone); characteristic heavy laryngealization (creaky voice for about 1/4 of duration at the end, alternatively glottal stop).

NV Stopped Drop Tone /.s/ Similar to the non-stopped drop tone, except for shorter duration (3/4 of drop tone, 3/5 of longer tones) and no laryngealization.

NV Curve Tone /’/ Average Fo level: low; concave Fo contour, starting low and falling to lowest Fo range at about 2/3 of duration, then rising toward the end (this rise varies according to speakers and phonetic environments); falling intensity contour; long duration; no characteristic laryngealization, but often breathy voice with speakers having low Fo ranges.

NV Broken Tone /~/ Average Fo level: mid to high; concave Fo contour, starting from mid Fo range, falling sharply to lowest Fo at about 1/3 of duration, then rising sharply again, reaching near top Fo range; concave intensity contour with drop corresponding to drop in Fo; long duration; characteristic mid-syllable laryngealization (creaky voice often heard as glottal stop).

(b) CV Tones

CV Level Tone /-/ Average Fo level: high (highest of CV tones); level Fo contour with a tendency to rise a little toward the end; falling intensity contour; long duration; no laryngealization.

CV Falling Tone /\/ Average Fo level: mid; falling Fo contour, starting above the mid Fo range and ending at low Fo level but not reaching the bottom range; falling intensity contour; long duration; no laryngealization.

CV Rising Tone /ʹ/ Average Fo level: mid; rising Fo contour, from a low level, with initial fall, then rising to high Fo but not reaching top range; sustained intensity contour; long duration; no characteristic laryngealization, but sometimes breathy voice in mid-syllable with speakers having low Fo ranges.

CV Stopped Rising Tone /ʹs/ Similar to the non-stopped rising tone but starting at higher Fo level and with short duration (about 3/5 of the rising tone).

CV Drop Tone /./ Average Fo level: low; falling Fo contour, starting below the mid Fo range, reaching lowest Fo at about 2/3 of duration then levelling out till the end; falling intensity contour; long duration; no characteristic laryngealization (except for some speakers who either have breathy voice or slightly creaky voice toward the end).

<u>CV Stopped Drop Tone /.s/</u> Falling Fo similar to the first 2/3 of drop tone, but starting a little higher; falling intensity contour; short duration (2/3 of drop tone); no laryngealization.

<u>CV Curve Tone /’/</u> Average Fo level: mid; concave Fo contour, with first 2/3 nearly coinciding with drop tone, last 1/3 rising (slope varies with speakers); falling intensity contour; slightly shorter duration (6/7 of drop tone); characteristic laryngealization (creaky voice ending or glottal stop).

(c) <u>SV Tones</u>

<u>SV Level Tone /-/</u> Average Fo level: mid; level Fo contour slightly falling toward the end; falling intensity contour; long duration; no laryngealization.

<u>SV Falling Tone /\`/</u> Average Fo level: low; falling Fo contour; falling intensity contour; long duration; no characteristic laryngealization, but sometimes breathy voice with speakers having low Fo ranges.

<u>SV Rising Tone /´/</u> Average Fo level: high; rising Fo contour, starting at about the same Fo as the level tone, and reaching top Fo range at the end; sustained intensity contour; long duration; no laryngealization.

<u>SV Stopped Rising Tone /´s/</u> Similar to the non-stopped rising tone in many respects, only the Fo onset is higher and the

duration shorter (3/5 of rising tone).

SV Drop Tone /./ Average Fo level: low; concave contour, starting at low Fo, reaching lowest Fo at about 1/3 of duration, then rising gently back to onset level; falling intensity contour; long duration; no characteristic laryngealization, but sometimes breathy voice with speakers having low Fo ranges, except for a local dialect (the Da Nang region) where creaky voice ending is often heard.

SV Stopped Drop Tone /.s/ Similar to the drop tone in many respects, but Fo onset a little higher, the rising end negligible, and shorter duration (3/5 of the drop tone).

SV Curve Tone /'/ Average Fo level: mid; concave Fo contour, starting at low level, reaching lowest Fo at about 1/3 of duration, then rising sharply but not reaching same height as the rising tone; sustained intensity contour, slightly concave; long duration; no characteristic laryngealization, but sometimes breathy voice with speakers having low Fo ranges.

From the foregoing description of individual tones, one can make the most general observations about the three dialect systems as follows.

Although the CV and SV systems differ from NV in the number of tones, there are more phonetic similarities between NV and SV tones than between these and the CV tones. Apart from intensity and duration which display no appreciable differences between the three dialects, the NV and SV systems have five phonetic tones which are similar in both Fo contour and average Fo level, namely the level tone, the falling tone, the rising tone, the stopped rising tone and the stopped drop tone. And they both differ from the CV tones in the use of Fo ranges: their tones reach more extreme Fo levels than their CV counterparts. These aspects will be discussed in more detail in the following subsections.

2.3.2 Fo Contour and Fo Level

Concerning these parameters, the following comments are relevant.

(a) The above description establishes four types of Fo contours in all three dialects : level, falling, rising and concave. As there is only one level tone in each dialect, it appears that Fo contour is the dominant feature in the Vietnamese tone system. This is further supported by the fact that all tones, with one exception, retain the same contour type across the three dialects and vary only in Fo level or laryngealization.

The level tone is mid-level in NV and SV, and high-level in CV;

The falling tone is low-falling in NV and SV, mid-falling in CV;

The rising and stopped rising tones are high-rising in NV and SV, mid-rising in CV;

The stopped drop tone is low-falling in all dialects;

The curve and broken tones are concave in all three dialects;

The exception is the drop tone, which is low-falling with laryngealized ending in NV, low-falling in CV and low-concave in SV.

These results accord with my auditory impressions and support the choice of contour features I presented in Ch. One (1.5.4.3).

(b) Average Fo levels can be determined more precisely by taking the arithmetic means of the Fo values at six timepoints of each tone (four timepoints for the stopped tones) from Tables 2.3 to 2.5 (pp 44-46). They are given in Table 2.11 (next page).

It appears from this table that three clear different Fo levels - high, mid and low - can be established for CV and SV tones, which also accords well with my auditory impressions. There is some ambiguity concerning the NV system: one could choose between three levels as in CV and SV, or only two levels (high, low) as there is, relatively, no great distance between the average Fo of the high and mid levels. Furthermore, from auditory impressions, I would classify the NV broken tone as high, probably because it has high ending Fo and the mid-syllable drop is often heard as a glottal stop and not low Fo. Indeed, if we leave out the low Fo value (169 Hz) at P3 which lies within the laryngealized part, the mean value for the other five timepoints of the broken tone is 219 Hz, just about the same as for the NV rising tone (218 Hz). I shall propose some conversion formulae (Ch. Five) which can overcome this problem and establish three average Fo levels for NV, with the level tone classified as 'mid' and the broken tone as 'high'.

In any case, the data in all three dialects support the choice of the phonological features 'high' and 'low' to represent three levels [+ high], [- high, - low] and [+ low]. As pointed out before, the three levels of average pitch in phonological features should not be confused with the five levels of pitch in Chao's

TABLE 2.11 Average Fo in Hz and Average Level of NV, CV and SV Tones (Data from Tables 2.3, 2.4 and 2.5)

Tone	NV		CV		SV	
	Aver. Fo	Level	Aver. Fo	Level	Aver. Fo	Level
Level /-/	209	Mid	185	High	189	Mid
Falling /\/	166	Low	175	Mid	155	Low
Rising /'/	218	High	171	Mid	219	High
St Ris /'s/	242	High	170	Mid	223	High
Drop /./	174	Low	153	Low	158	Low
St Drop /.s/	170	Low	158	Low	161	Low
Curve /'/	160	Low	157	Low	183	Mid
Broken /~/	211	Mid				
Range of Aver. Fo for each level						
High	218-242		185		219-223	
Mid	209-211		170-175		183-189	
Low	160-174		153-158		155-161	

system which I also use to record details of onset, midpoint and endpoint pitch targets for each tone, e.g. [35], [214] in systematic phonetic representations.

Concerning the stopped tones, the data show that, notwithstanding differences in detail at onset and endpoint, they have similar Fo contour and average Fo level of the corresponding non-stopped tones, as well as similar intensity contour. I shall, in Ch. Three (3.1.2), show that the differences in duration and other phonetic details may be intrinsically conditioned by the presence of the final voiceless stop, and therefore predictable.

2.3.3 *Fo Ranges*

The physiologically caused intrinsic differences in Fo ranges between male and female speakers are well known, but one may suspect that other factors are involved in determining the actual Fo ranges used by speakers within the possible limits.

The absolute and mean Fo ranges of individual NV, CV and SV informants are presented in Table 2.12 and Figures 2.10 to 2.12 (following pages). The figures in Table 2.12 are calculated according to two similar formulae for the absolute range R and the mean range $\bar{R}$:

$$R = Fmax - Fmin$$

$$\bar{R} = \bar{F}max - \bar{F}min$$

where the range R was calculated from the highest absolute value Fmax and the lowest absolute value Fmin, and the mean range $\bar{R}$ was calculated from the highest mean value $\bar{F}max$ and the lowest mean value $\bar{F}min$, for each individual informant. The largest and smallest absolute and mean ranges by individual informants of the three dialects are as follows.

	NV	CV	SV
Largest R	230 Hz (NF4)	130 Hz (CM1)	240 Hz (SF3)
Largest $\bar{R}$	167 Hz (NF2)	73 Hz (CM7)	173 Hz (SF5)
Smallest R	125 Hz (NM4)	60 Hz (four)	85 Hz (SM4, SM6)
Smallest $\bar{R}$	94 Hz (NM4)	32 Hz (CM5)	59 Hz (SM2)

The mean Fo ranges and standard deviations of the three groups of informants are

	n	$\bar{R}$	SD
NV	9	131	24
CV	12	54	13
SV	9	115	45

Results of t-tests show that there is no significant difference between the mean ranges of the NV and SV groups, but significant differences at 0.01 level between those of the CV group and each of the other groups. (The t-test formula is given on p. 87.)

The data suggest that the smaller Fo ranges used by CV speakers are not due to individual characteristics but rather to the characteristics of the CV tone system itself, probably the fact that CV has no high rising but only mid rising tones, and the highest Fo reached by its high level and mid rising tones cannot be so high as those reached by the NV and SV high rising tones.

2.3.4 Intensity

The general impression one gets from the intensity curves plotted in Figures 2.7 (a,b), 2.8 (a,b) and 2.9 (a,b), on pages

TABLE 2.12 Absolute and Mean Fo Ranges of Individual Informants

Informant	Fmax	- Fmin	= R	$\overline{F}$max	- $\overline{F}$min	= $\overline{R}$ *
NF2	380	170	210	354	187	167
NF4	370	140	230	314	157	157
NF3	350	170	180	329	176	153
NF1	310	170	140	298	179	119
NM3	310	140	170	291	146	145
NM6	270	140	130	261	144	117
NM1	250	90	160	199	98	101
NM2	240	80	160	217	91	126
NM4	205	80	125	184	90	94
CF2	300	200	100	276	206	70
CF1	280	185	95	262	193	69
CF4	280	160	120	232	166	66
CF3	275	205	70	259	210	49
CM1	270	140	130	211	148	63
CM7	240	130	110	212	139	73
CM8	190	110	80	179	123	56
CM4	165	105	60	157	118	39
CM2	165	100	65	157	103	54
CM5	165	105	60	146	114	32
CM6	160	95	60	140	101	39
CM3	150	90	60	137	95	42
SF1	400	230	170	391	252	139
SF5	400	180	220	366	193	173
SF3	400	160	240	336	173	163
SF2	360	125	235	299	133	166
SM1	280	150	130	257	156	101
SM5	220	100	120	204	107	97
SM2	190	100	90	163	104	59
SM4	180	95	85	168	99	69
SM6	175	90	85	162	92	70

* Values in Hz. For calculation formula see text (previous page).

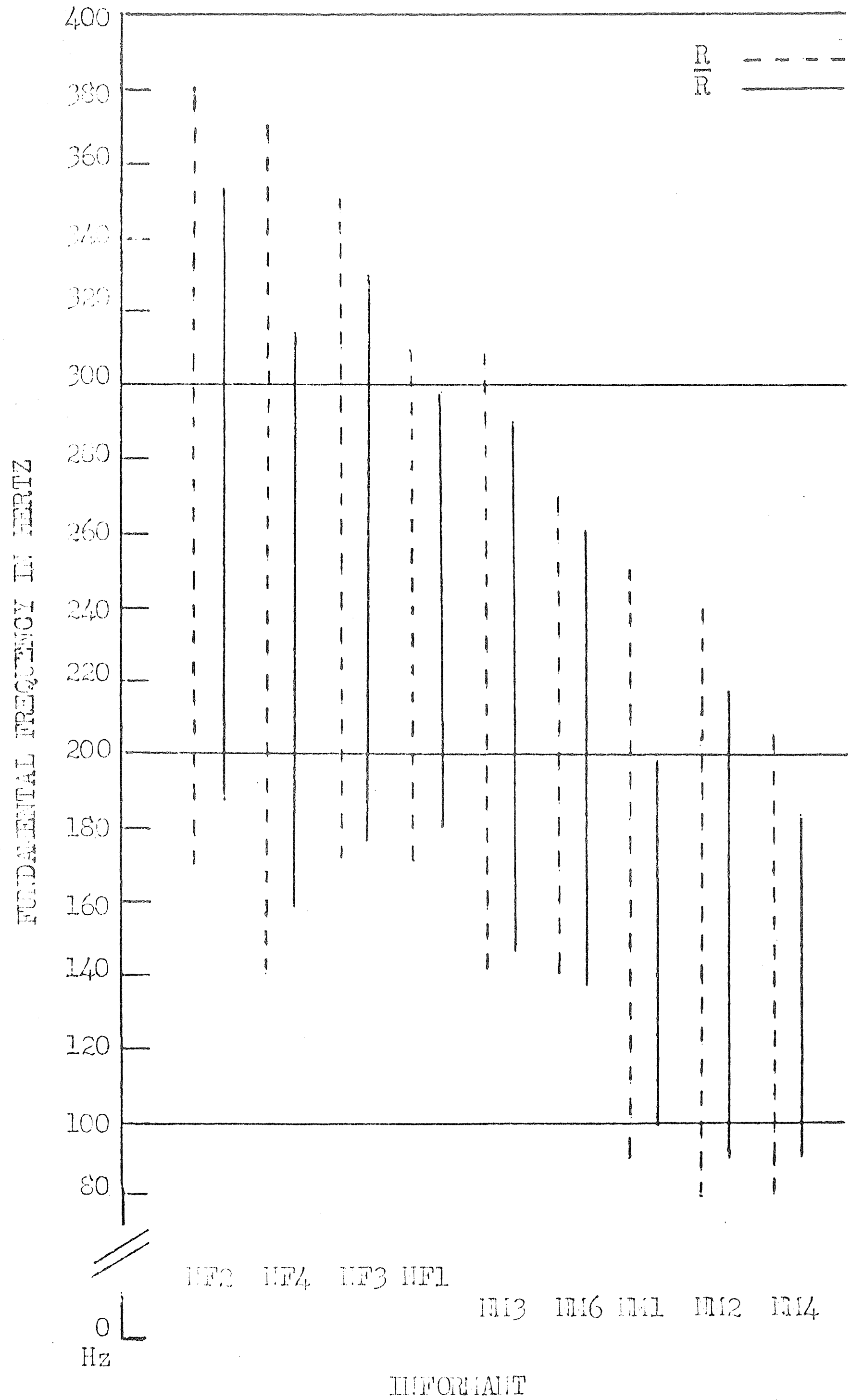


FIGURE 2.10 Absolute (R) and Mean ($\bar{R}$) Fo Ranges of Individual NV Informants (ordered in decreasing highest absolute Fo).

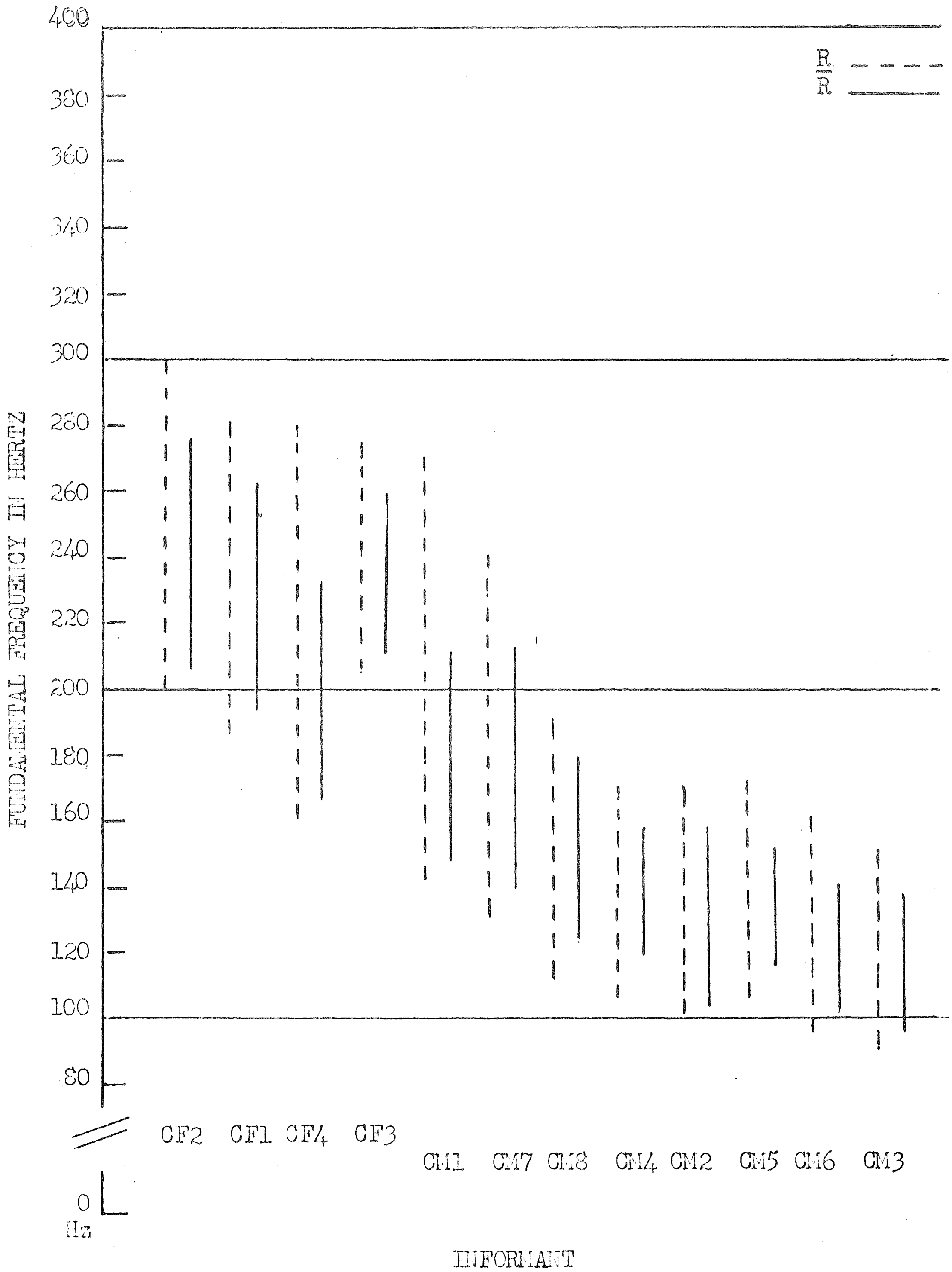


FIGURE 2.11 Absolute (R) and Mean ($\bar{R}$) Fo Ranges of Individual CV Informants (ordered in decreasing highest absolute Fo).

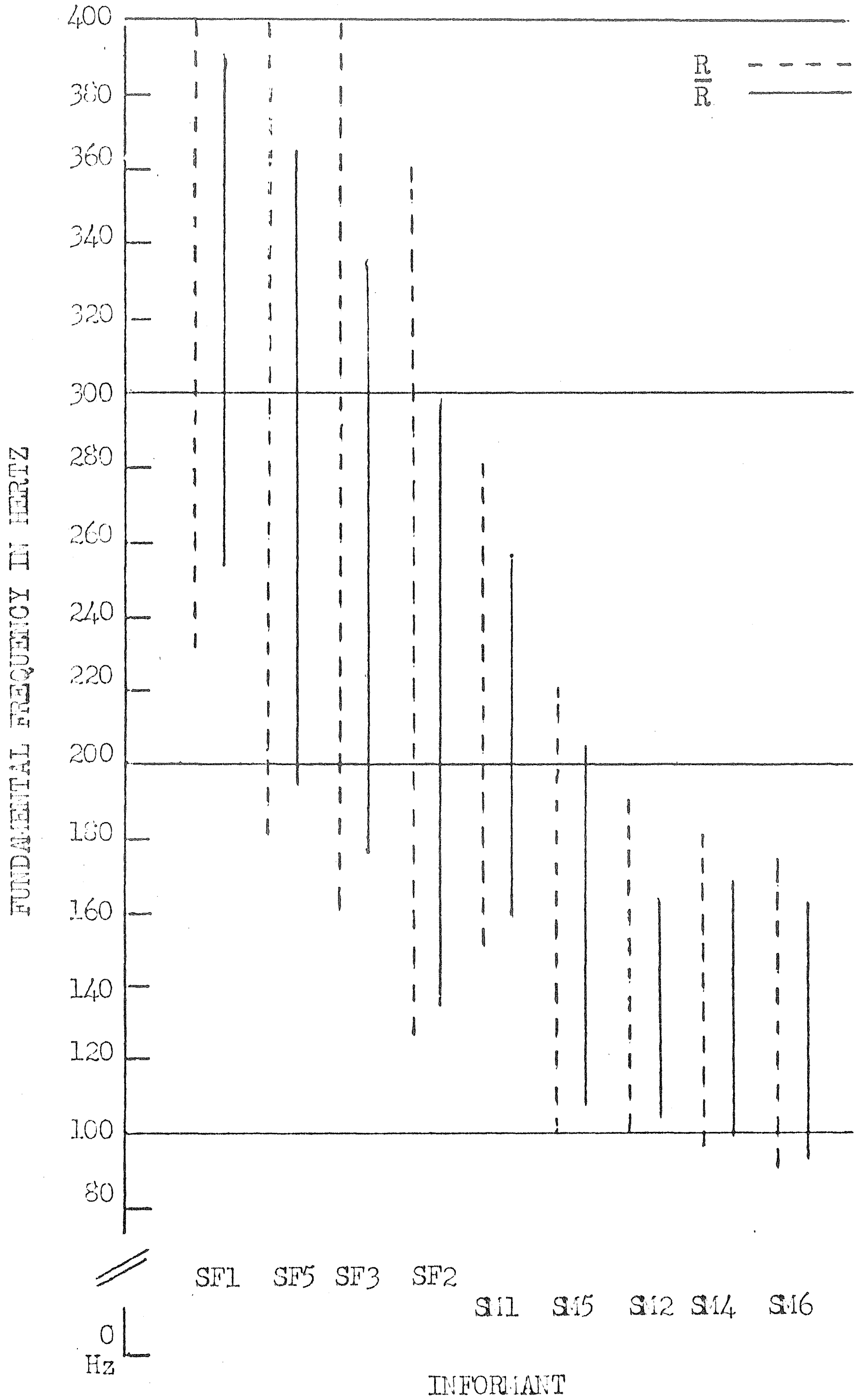


FIGURE 2.12 Absolute (R) and Mean ($\bar{R}$) Fo Ranges of Individual SV Informants (ordered in decreasing highest absolute Fo).

53-58, is that intensity covaries with Fo to a certain degree only. Most intensity curves have falling slopes in tones with level or falling Fo contours, while some have a more sustained or slightly rising slope in the middle of a rising tone, but start to decay before Fo reaches its peak toward the end. There seems to be little difference between dialects and individuals. This is borne out in Table 2.13 (next page) where the correlation coefficient r between intensity and Fo values at four timepoints A, B, C and D (representing I1, I2, I3 and I4 for intensity and the corresponding P2, P3, P4 and P5 for Fo) was calculated for each timepoint from the data by six different informants from the three dialects. Most of the values of r vary between .27 and .69, that is they show a fair degree of correlation between the two sets of data. The exceptions are a fairly high correlation (.75) at timepoint D for Informant SF5 and no correlation (.03) at the same timepoint for Informant SM4 .

This suggests that in this respect Vietnamese may be different from some other tone languages such as Mandarin Chinese for which Kratochvil 1968 has established intensity as an important characteristic of tone, or the Ningpo Chinese dialects for which Rose 1981 has shown that intensity "varies independently of Fo" and "varies regularly as a characteristic of a given lexical tone."

However, I shall show below that there is a possibility that in some cases, intensity contour may become a major cue for identifying the NV broken tone.

Furthermore, it has been shown that intensity may vary with different vowels (Lehiste 1970) and I have observed that it

TABLE 2.13 Correlation between Fo and Intensity at Four Timepoints of Same Syllables with NV, CV and SV Tones by Six Individual Informants

Informant	n	rA	rB	rC	rD *
NF2	32	.53	.53	.49	.36
NM6	32	.27	.30	.39	.35
CF1	28	.49	.68	.69	.69
CM7	28	.39	.39	.61	.27
SF5	28	.53	.29	.60	.75
SM4	28	.53	.31	.44	.03

* A, B, C and D represent corresponding data points for Fo and intensity; r is the product-moment correlation coefficient calculated according to the statistical formula

$$r = \left\{ \frac{1}{n} \left(\sum_{i=1}^{n} x_i y_i \right) - \overline{x}\overline{y} \right\} / S_x S_y$$

where

$x_i \ldots\ldots x_n$ and $y_i \ldots\ldots y_n$ are the two sets of data, $\overline{x}$ and $\overline{y}$ are the respective means of the sets, and S_x and S_y are their respective standard deviations.

usually rises slightly toward the end of syllables with a sonorant final consonant. Figure 2.13 (next page) illustrates some of the above mentioned points.

2.3.5 Duration

There is considerable difference in absolute values for the duration of individual syllables by different informants, but the mean values show that in each dialect, the stopped tones are all shorter than the average non-stopped tones by about 1/3. The two crealy-ending tones in NV and CV are shorter by about 1/5 and 1/7 respectively than the other non-stopped tones. I shall deal with the phenomenon in more detail later, in 2.4.3.4 infra, after normalization makes interdialect comparison more meaningful.

2.3.6 Laryngealization

Some useful observations can be made about the occurrence of laryngealization in the three dialects as summarized in Table 2.10 (p 51).

(a) Besides occurring as a regular feature on the NV drop and broken tones and the CV curve tone, creaky voice also occurs on a local variant of the SV drop tone in the Da Nang region (the northernmost part of the SV dialect area) and of the drop tone with some CV informants from the Vinh region (the northernmost part of the CV dialect area), albeit with a lesser degree of laryngealization, so far as I can judge from my ears. In the latter case, CV speakers from the region might tend to emphasize the rising end of the curve tone to make it more distinct from the drop tone, or to confuse the two tones together and have one tone less

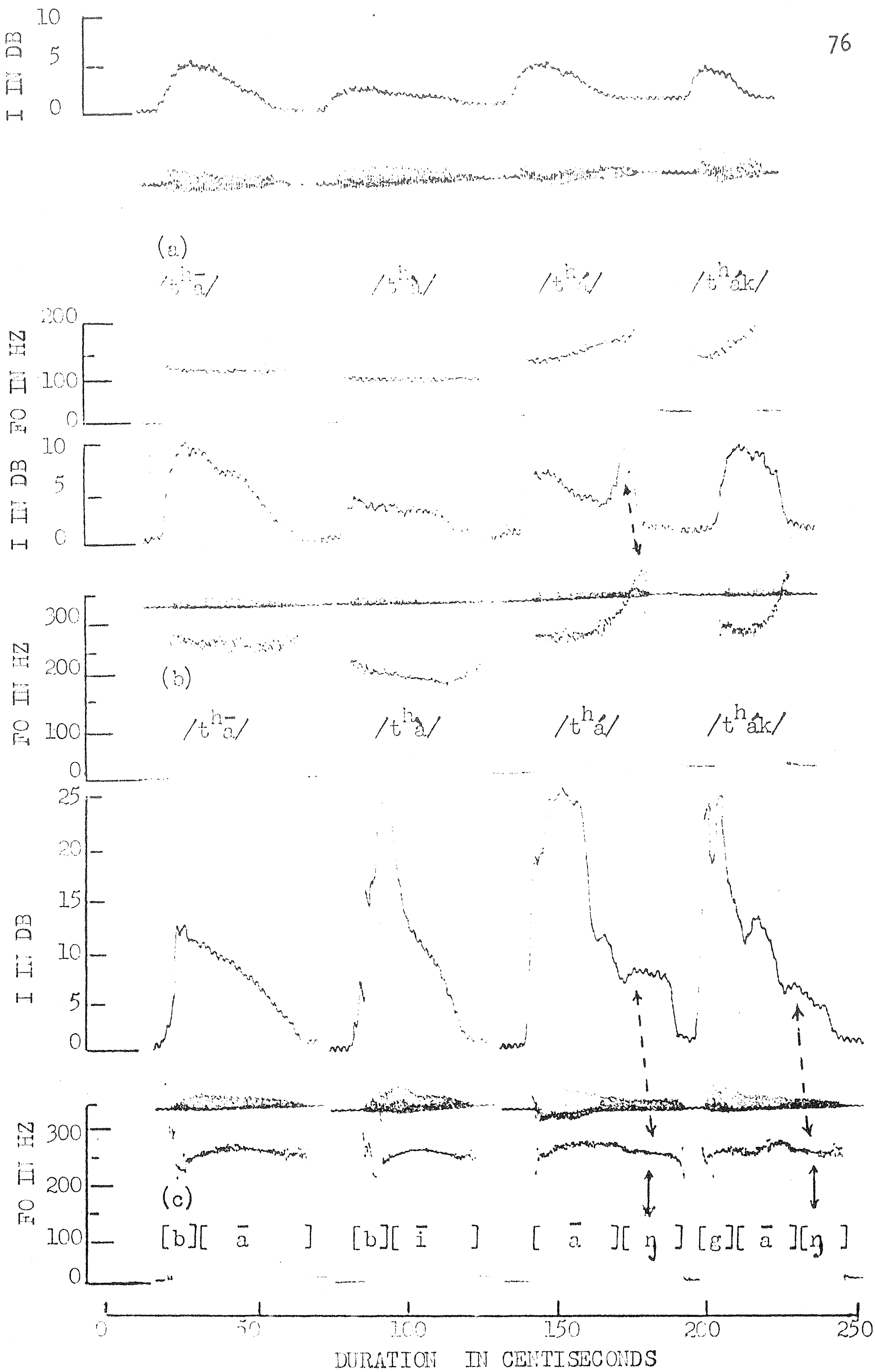


FIGURE 2.13 Intensity and Fo Curves of SV Tones by Informants (a) SM4 and (b)(c) SF5 (Arrows suggesting correlation between Fo and I and consonant and I are slightly slanted because of a difference in integration time between the Fo and I curves)

in their system. This might be one of the causes for tone changes across Vietnamese dialects.

(b) Creaky voice is not always realized in the same way: it varies greatly in duration between informants and is sometimes replaced by glottal closure, which is the reason why many authors described it as a glottal stop. When it occurs in the middle of the broken tone, the syllable is sometimes heard as two, as reported in Vu Ba Hung 1978, by foreign listeners.

It is interesting to see how the same auditory effect could be produced in different ways by examining some of the mingograms and spectrograms of the NV broken tone on the same syllable /tã/ by three different informants, NF3, NM3 and NM4 (Figure 2.14, next page). These speakers' tone system does not show any noticeable deviations from standard NV and their creaky tones produce similar auditory impressions. However, the voiceprints of the same syllables, which are typical of each informant, present marked differences:

NF3 /tã/ : heavy laryngealization, sharp drop in Fo and in intensity at middle;

NM3 /tã/ : no laryngealization, sharp drop in Fo and slight drop in intensity at middle;

NM4 /tã/ : no laryngealization, no sharp drop in Fo but only sharp drop in intensity at middle. For comparison, the mingogram of the syllable with rising tone /tá/ by the same informant NM4 is included alongside /tã/: it is clear that the Fo curves of the two tones by this informant are similar, only the intensity contours differ sharply.

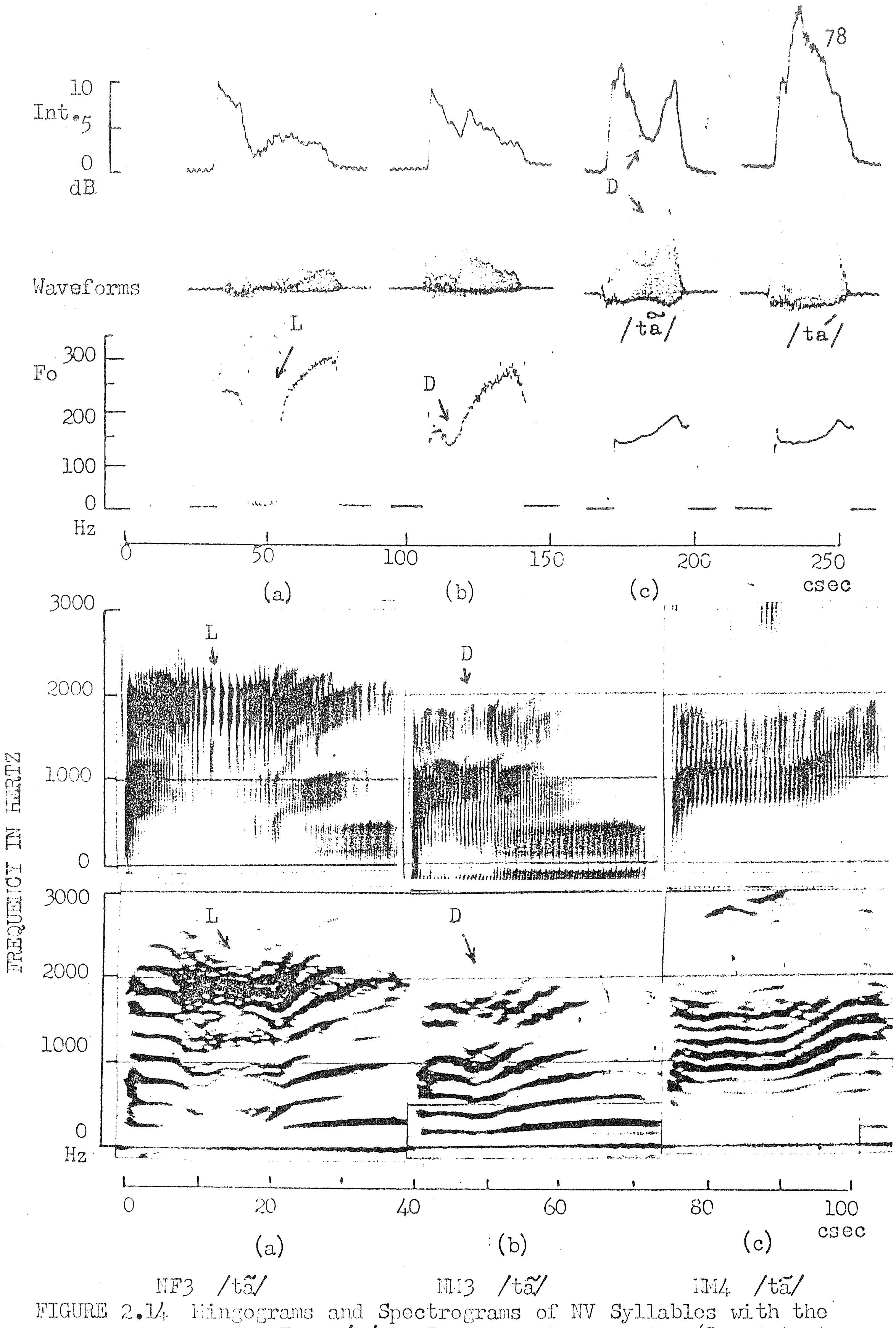


FIGURE 2.14 Mingograms and Spectrograms of NV Syllables with the Broken Tone /~/ by Different Informants. (L points to laryngealization; D points to a characteristic drop; mingograms of the syllable /tá/ by Informant NM4 are included for comparison with /tã/)

(c) Breathy voice, which occurs irregularly in all three dialects, is usually associated with low tones and particularly with speakers having lower Fo ranges and can manifest itself in different degrees. A spectrogram of a syllable with typical breathy voice by Informant NM4 (whose NV falling tone and curve tone are regularly breathy almost through their lengths) is shown on the upper right corner of Figure 2.15 (next page) beside a syllable with the characteristic creaky voice ending by Informant CF3. Both syllables bear the same tone /'/ with different shapes in the two dialects. The lower part of Figure 2.15 shows three successive occurrences of the stopped rising tone on three syllables by the same Informant CF3, who usually has slightly creaky voice in the middle of her rising tones on some syllables. These are interesting things about the syllables in the figure. The first syllable /t^hák/ shows slight laryngealization for some duration in its middle; the auditory effect of this is very slight creaky voice that can be heard, but which would not be noticed in speech because it would not be expected there. The two other syllables are quite clear on hearing, but the middle syllable /dák/ displays some very short perturbations in the middle which are not ordinarily audible but might be a potential cause for the development of laryngealization in certain circumstances.

2.4 *PHYSICAL PHONETIC PARAMETERS OF VIETNAMESE TONES*

The discussion of results of direct measurements of the acoustic data in absolute or mean values has provided us with a

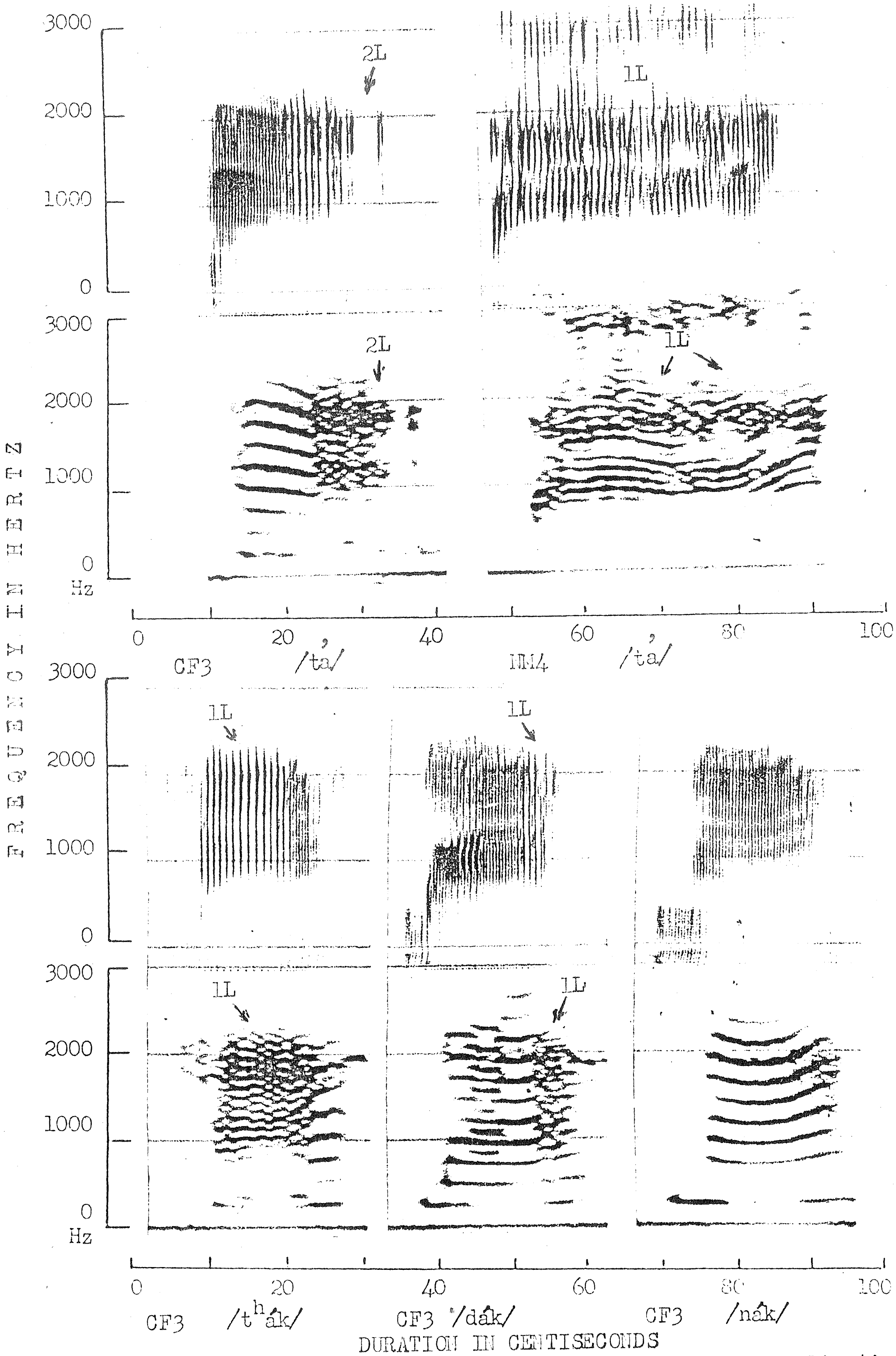


FIGURE 2.15 Spectrograms Showing Varying Degrees of Laryngealization (1L : slight; 2L : heavy) in CV and NV Syllables

Note that the laryngealization indexes are based on my auditory impression.

fair idea of the characteristics of NV, CV and SV tones in their respective systems. However, the values given in these results also represent some variations in non-linguistic parameters such as Fo range differences between speakers or groups of speakers, differences in their speech tempo or the level of their voices, etc, which make the data not directly comparable.

The normalization procedures proposed below serve a double purpose. First, they will make any two or more sets of data from different speakers or groups of speakers directly comparable by practically eliminating non-linguistic variables and presenting the normalized values in easily interpretable percentage or decimal scales, as exemplified in 2.0. Second, they will provide these normalized values as input for the conversion processes described later in Ch. Five, as part of a proposed model of tone perception.

2.4.1 Normalization Procedures

As proposed earlier, the four sets of parameters needed to describe Vietnamese tones adequately at the physical phonetic level are fundamental frequency Fo, intensity I, duration D and laryngealization L. Laryngealization has been described in 2.3.6 above and needs no normalization. The other data concerning Fo, I and D can usually be normalized in percentage or decimal scales.

2.4.1.1 Fo Differential in Percent A usual procedure for normalization of Fo data is a percentage scale based on the average range, as that used by Earle 1975 for NV tones. It is adequate for a fairly homogeneous tone system. However, when looking at the Fo data for the three Vietnamese dialects, one is struck by the fact that the CV informant group made use of a much smaller Fo range than the other two groups, as noted earlier. A percentage scale based on the average range would exaggerate the Fo slopes of CV tones as compared with the NV and SV systems (as I shall show when discussing the results of normalization in 2.4.3 below.) Therefore, for comparison of Fo in different situation, I devised a method of normalization involving the concept of Fo Differential or FD, expressed in the following formulae

$$FD(A,B) = Itg\left(\frac{F_i(B) - F_i(A)}{F_i(A)} \times 100\right) \qquad (N1)$$

$$FD(\bar{F}) = Itg\left(\frac{F_i - \bar{F}}{\bar{F}} \times 100\right) \qquad (N2)$$

where F_i is any individual Fo value,

A and B are any two timepoints at which F_i occurs,

$\bar{F}$ is the mean Fo of a sample, used as a reference level,

and Itg stands for integer, i.e. the FD will be expressed in integer digits, any decimals being automatically cut off.

The FD values can be positive or negative and represent the differential percentage (higher or lower, rise or fall) between two individual Fo values or between an Fo value and the mean $\bar{F}$.

Formula (N1) calculates the FD between point A and point B from the Fo values of those points, $F_i(A)$ and $F_i(B)$, and gives the differential percentage in relation to point A. A positive value means that F_i at point B is higher than at point A, and a negative value means that F_i at point B is lower than at point A. This formula can be used to calculate the FD between two points on the same Fo curve, e.g. the onset and endpoint of the same tone. The resulting FD, say -12, indicates a fall of 12% in Fo between these two points. Or it can be used to calculate the FD between two comparable points on different Fo curves, e.g. the onsets of two tones on the same syllables. The resulting FD, say 8, indicates that there is a difference of 8% in Fo onset between the two tones, in relation to the lower onset.

Formula (N2) calculates the FD between any individual F_i and the mean $\bar{F}$ for any informant or group of informants. The mean $\bar{F}$ is the arithmetical mean of all Fo values of a sample (of p time-points x s syllables x t tones x n speakers). The mean $\bar{F}$ used for calculating the Fo parameters in FD($\bar{F}$) percent of NV, CV and SV tones presented in Table 2.15 infra, obtained from the total sample for each dialect, were respectively 193 Hz, 164 Hz and 183 Hz for the three groups of NV, CV and SV informants. Thus if a tone has an FD($\bar{F}$) of -6 at its onset and -15 at its endpoint, for example, it can be inferred from the negative values that it is a low and falling tone, irrespective of the actual Fo values.

These are some examples of calculations.

(a) The differences between mean Fo onset and endpoint values for the falling tone are respectively:

174 - 156 = 18 Hz for the CV group of informants, and

166 - 149 = 17 Hz for the SV group.

If we apply formula (N1), in which 0 stands for onset and E for endpoint

$$FD(0,E) = \text{Itg}\left(\frac{F_i(E) - F_i(0)}{F_i(0)} \times 100\right)$$

we shall have, for the CV falling tone

$$FD(O,E) = \text{Itg}\left(\frac{156 - 174}{174} \times 100\right) = -10\%$$

and for the SV falling tone

$$FD(O,E) = \text{Itg}\left(\frac{149 - 166}{166} \times 100\right) = -10\%$$

Thus, in spite of differences in absolute values, the CV and SV falling tones can be said to have the same rate of fall, -10% between onset and endpoint Fo. This would mean that, heard out of context, they would be both identified as a falling tone of the same phonetic shape.

(b) If we apply formula (N2) to calculate the FD relative to the mean $\bar{F}$

$$FD(\bar{F}) = \text{Itg}\left(\frac{F_i - \bar{F}}{\bar{F}} \times 100\right)$$

in which $\bar{F}$ is 164 Hz for the CV group and 183 Hz for the SV group of informants, we shall have the following values for the same CV falling tone

$$FD(\bar{F}) \text{ at onset} = \text{Itg}\left(\frac{174 - 164}{164} \times 100\right) = +6\%$$

$$FD(\bar{F}) \text{ at endpoint} = \text{Itg}\left(\frac{156 - 164}{164} \times 100\right) = -4\%$$

and for the SV falling tone

$$FD(\bar{F}) \text{ at onset} = \text{Itg}\left(\frac{166 - 183}{183} \times 100\right) = -9\%$$

$$FD(\bar{F}) \text{ at endpoint} = \text{ltg}\left(\frac{149 - 183}{183} \times 100\right) = -18\%$$

Thus, although these two falling tones have the same slope and no great differences in mean Fo values, the FD($\bar{F}$) values clearly establish that the CV falling tone is mid-falling (having an onset 6% above and an endpoint -4% below the mean $\bar{F}$), and that the SV falling tone is low-falling (both onset and endpoint are well below the mean $\bar{F}$, viz. -9% and -18%). This would mean that, heard in tonal context in their own systems, the CV and SV falling tones would be rated differently in their average pitch levels. This property of the FD($\bar{F}$) to suggest both the Fo contour and the Fo level relative to the mean $\bar{F}$ makes it a useful tool for inter-dialect and inter-speaker comparisons and also for modelling the perceptual process, as I suggested earlier. That is why I have adopted it in preference to other possible normalization procedures.

Note that for convenient reference, the FD($\bar{F}$), called Fo Differential relative to the mean $\bar{F}$ for any informant or group of informants, will be always written in this form, whereas the Fo Differential between two points A and B can be indifferently referred to as FD(A,B) or 'the FD between A and B in relation to A' (with A being the basis for percentage calculation). Note also that the - sign is mandatory, while the + sign is usually omitted.

The relative advantage of this method over the percentage-of-range scale will be apparent in comparing the Fo values obtained from different informants who may have different pitch levels and voice ranges due to their sex or age, or local dialect, or the conditions of recording at different times. Any differences can thus be expressed in percentage instead of absolute values,and the point of reference, the mean fundamental frequency, is not a fixed value but a movable one calculable from any particular set of variables from any population sample.

2.4.1.2 Decimal Scale for Intensity and Duration For the I and D parameters, I propose to use a decimal scale for representing the relative intensity contour and the relative duration of each tone in each dialect system. This scale is chosen for two reasons: like the percentage scale, it is relatively easy to convert actual values to a decimal scale by simple mathematical formulae; second, it is probable that differences of at least one-tenth or more in relative intensity or duration are auditorily detectable or statistically significant. Table 2.14 (next page) presents some levels of significance obtained from t-tests between the means of some intensity and duration data. The results reported in this table suggest, for instance, that for some CV tones, intensity differences of 0.8 db (less than 10% of the mean peak intensity in the sample) are not significant at the 0.5 level of confidence, but differences of 0.9 dB (about 11% of the mean peak intensity) are significant at that level. Similarly, differences of two centiseconds in the mean duration are not statistically significant at the 0.5 level of confidence, but those of four or five centiseconds are found to be highly significant, i.e. at the 0.1 or 0.2 level.

To convert actual values to the decimal scale, for both intensity and duration, the following normalization formulae were applied:

TABLE 2.14 Levels of Significance Obtained from t-Tests On Some Intensity and Duration Data*

Data Tested	n1	n2	$\bar{x}1$	$\bar{x}2$	S1	S2	t	p **	Conclusion
$\bar{I}2$ & $\bar{I}3$ of NV /-/	32	32	6.3	5.4	1.1	1.6	2.21	<0.05	significant
$\bar{I}2$ & $\bar{I}1$ of SV /-/	32	32	5.8	4.9	1.4	2.0	2.05	<0.05	"
$\bar{I}1$ & $\bar{I}2$ of CV /-/	32	32	8.0	7.2	2.5	1.1	1.63	>0.05	not signif.
$\bar{I}1$ of CV /ˋ/ & $\bar{I}_1$ of CV /'/	32	32	8.4	7.3	2.7	2.7	2.10	<0.05	significant
$\bar{D}$ of NV /-/ & NV /./	8	8	25	20	6	4	4.6	<0.01	significant
$\bar{D}$ of CV /-/ & CV /ˋ/	8	8	26	28	5	5	1.57	>0.05	not signif.
$\bar{D}$ of CV /./ & CV /'/	8	8	28	24	7	5	2.58	<0.02	significant
$\bar{D}$ of SV /-/ & SV /'/	8	8	30	28	8	6	0.52	>0.05	not signif.

* Data from Tables 2.6 to 2.9.

** The statistical formula for the two-tailed t-test is

$$t = \frac{|\bar{x}1 - \bar{x}2|}{\sqrt{\frac{(n1 x S_1^2) + (n2 x S_2^2)}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the means of sample 1 and sample 2 respectively,

n_1 and n_2 are the number of tokens in those samples,

S_1 and S_2 are the respective standard deviations, and

the probability p reported for a critical value of t in each test indicates the level at which there is 95% ($p<0.05$) or 99% ($p<0.01$) probability that the differences between two sets of data are statistically significant, i.e. not due to chance.

$$I = \mathrm{Itg}\left(\frac{I_i \times 10}{\bar{I}max} + 0.9\right) \qquad \text{(N3)}$$

$$D = \mathrm{Itg}\left(\frac{D_i \times 10}{\bar{D}max} + 0.9\right) \qquad \text{(N4)}$$

where I_i and D_i are any I and D values to be normalized,

$\bar{I}max$ and $\bar{D}max$ are the highest mean values in the samples in question (i.e. $\bar{I}max$ is the highest mean intensity for all tones in a sample at a particular timepoint , and $\bar{D}max$ is the longest mean duration for any tone in the sample.)

As in the case with Fo, I and D are expressed in integer digits, and to ensure that the decimal scale of 1 to 10 corresponds to the range of 0.01 - 0.1 for 1, up to 0.91 - 1.0 for 10, a correcting factor of 0.9 is added in the formula. Thus the mean value at I2 for the intensity of the NV level tone,has a normalized value of

$$\mathrm{Itg}\left(\frac{6.3 \times 10}{7.1} + 0.9\right) = \mathrm{Itg}(9.7) = 9$$

and the mean duration value for the SV stopped drop tone has a normalized value of

$$\mathrm{Itg}\left(\frac{17 \times 10}{30} + 0.9\right) = \mathrm{Itg}(6.5) = 6$$

Note that all those percent and decimal scales can be conceived as open-ended scales useful for the description of extreme values of individual occurrences of some tones in comparison with the standard forms.

2.4.2 Physical Phonetic Parameters of NV, CV and SV Tones in Normalized Values

The normalization formulae and the four degrees of laryngealization proposed earlier are useful tools for determining tone parameters in different dialects in a uniform way. They are given in Table 2.15 (next page).

Graphic representations of these parameters are given in Figures 2.16 and 2.17 (following pages).

In Figure 2.16 the tones are represented in a percent-of-average-range scale for Fo and a decimal scale for duration. The percentage scale, as used in Earle 1975, involves the determination of the average range $\bar{R}$ for the three groups of NV, CV and SV speakers and the conversion of Fo values to percent-of-average-range values by the formulae

$$\bar{R} = \bar{F}max - \bar{F}min \qquad (N5)$$

$$F_r = \frac{F_i - \bar{F}min}{\bar{R}} \times 100 \qquad (N6)$$

where $\bar{R}$ is the average range,

$\bar{F}max$ and $\bar{F}min$ are the highest and lowest mean values in the sample, F_r the Fo value in percent of range and F_i the actual value to be converted.

Figure 2.17 represents the tones in their FD($\bar{F}$) percent values and mean duration in decimal scale. Note that while the percent-of-average range correctly represents the relative Fo contours and pitch levels within a dialect, it distorts the slopes of Fo contours of CV tones in comparison with the other dialects. The

TABLE 2.15 Physical Phonetic Parameters (Fo, L, I & D in Normalized Values) of NV, CV and SV Tones *

Dial. & Tone	Fo in FD($\bar{F}$) Percent						Laryng.			Intensity				Duration **
	P1	P2	P3	P4	P5	P6	a	b	c	I1	I2	I3	I4	
NV /-/	9	9	9	8	7	4	0			10	9	8	3	10
/ˎ/	-7	-9	-10	-15	-17	-19	0			8	7	5	2	10
/ʹ/	2	0	2	12	27	33	0			8	7	8	3	10
/ʹs/	14	20	29	38			0			8	7	7	3	6
/./	-2	-3	-6	-14	-16	-15	2	2	E	7	6	4	2	8
/.s/	-5	-9	-13	-17			0			7	7	5	2	6
/ʼ/	-8	-11	-18	-26	-19	-13	0			7	5	4	2	10
/~/	4	1	-12	9	26	26	2	2	M	8	4	7	2	10
CV /-/	12	10	12	14	14	13	0			10	9	7	4	10
/ˎ/	6	4	1	-1	-3	-4	0			10	8	5	3	10
/ʹ/	-3	-6	-8	-1	7	10	0			7	6	6	3	10
/ʹs/	0	-3	1	9			0			7	6	6	2	6
/./	0	-3	-6	-9	-9	-9	0			8	6	4	2	10
/.s/	3	-1	-5	-8			0			8	7	5	2	6
/ʼ/	1	0	-4	-9	-6	-4	2	2	E	9	7	4	2	9
SV /-/	4	4	4	4	3	1	0			9	10	7	4	10
/ˎ/	-9	-12	-14	-17	-18	-18	0			6	7	5	2	10
/ʹ/	6	4	9	22	36	39	0			8	9	8	5	10
/ʹs/	9	13	24	39			0			8	10	9	5	6
/./	-9	-14	-19	-18	-11	-9	0			6	5	4	2	10
/.s/	-7	-13	-15	-12			0			7	7	5	2	6
/ʼ/	-5	-12	-18	-4	19	22	0			7	5	6	4	10

* Calculated from data in Tables 2.3 to 2.10.

** In the Laryngealization parameter, (a) indicates degree, (b) duration (same scale as for whole tone), and (c) timing. Intensity and duration values are in decimal scales. For more detail see 2.4.1.

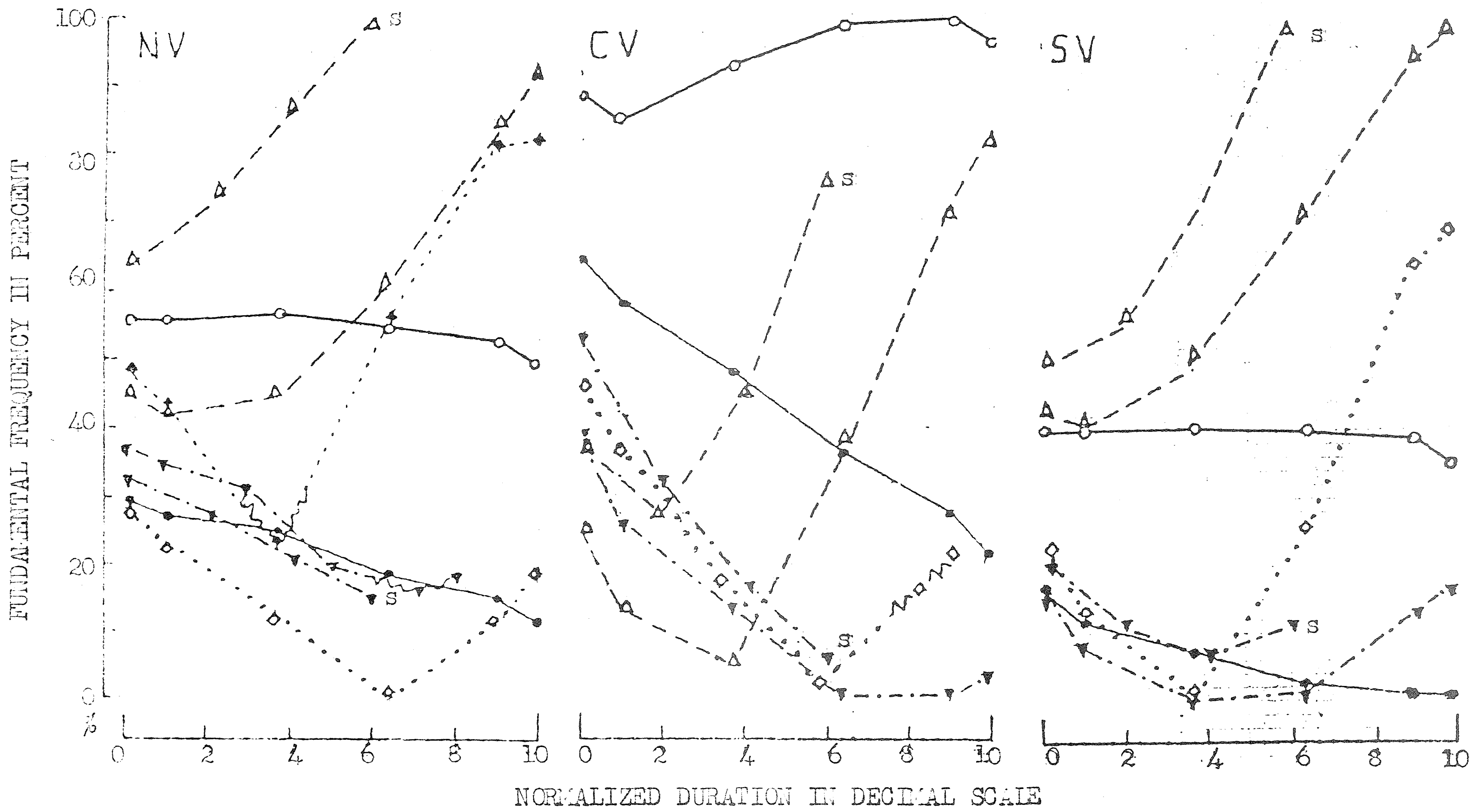
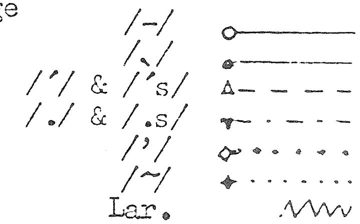


FIGURE 2.16 Mean Fo of NV, CV and SV Tones Converted to a Percent-of-Average-Range Scale Plotted Against Normalized Duration



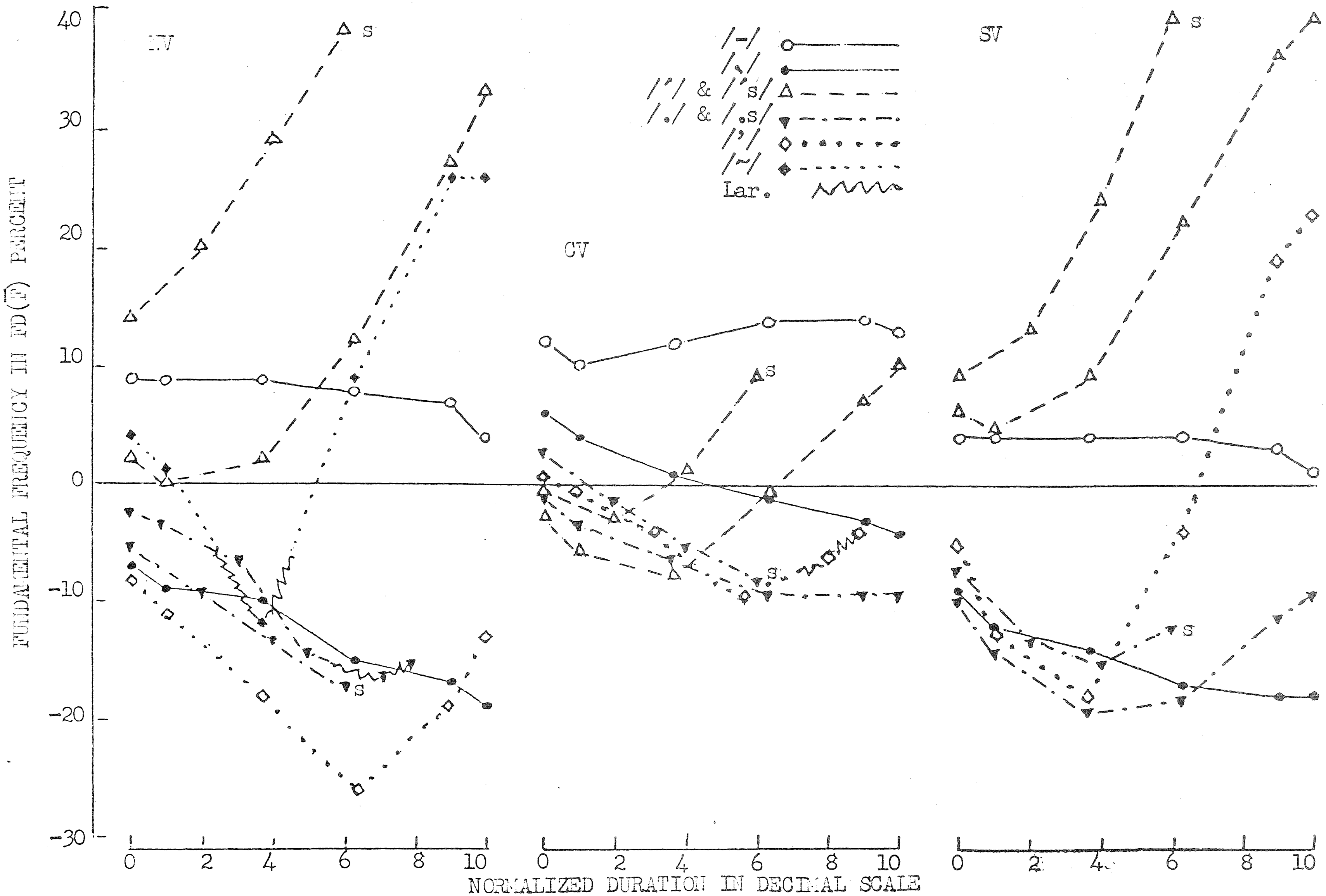


FIGURE 2.17 Normalized Mean Fo in FD(F̄) Percent of NV, CV and SV Tones Plotted Against Normalized Duration

$FD(\bar{F})$ scale preserves the basic shapes of Fo contours in the three dialects but brings the Fo levels into more comparable perspectives.

Figure 2.18 (next page) presents the normalized intensity contours plotted against normalized duration.

2.4.3 Discussion of the Normalized Parameters

2.4.3.1 Fo in $FD(\bar{F})$ Percent The $FD(\bar{F})$ values, as given in Table 2.15 and plotted in Figure 2.17 above, give a correct representation of the main characteristics of the NV, CV and SV tones individually and within their systems, as discussed in 2.3.1 and 2.3.2 supra. In addition to the example I gave in 2.4.1 concerning the falling tone, the study of the typical onset, midpoint and endpoint $FD(\bar{F})$ values of any tone in Table 2.15 would enable us to determine its characteristic Fo contour and Fo level in a rough manner. The NV and SV non-stopped and stopped rising tones have all positive values with very high endpoints (up to between 33% and 39%) are clearly high-rising, while their CV counterparts, having negative values in the first half and positive values in the latter half, are clearly mid-rising.

The highest and lowest $FD(\bar{F})$ values in the three dialects (NV : +38, -26; CV : +14, -9; SV : +39, -19) also reflect the typically narrower Fo range used by CV speakers as observed in 2.3.3.

2.4.3.2 Variation in Fo Contour and Slope The normalization formulae allow us to compare not only the tones in different dialects, but also those between individual aspeakers with different voice ranges. Examples of variation in Fo contour and slope

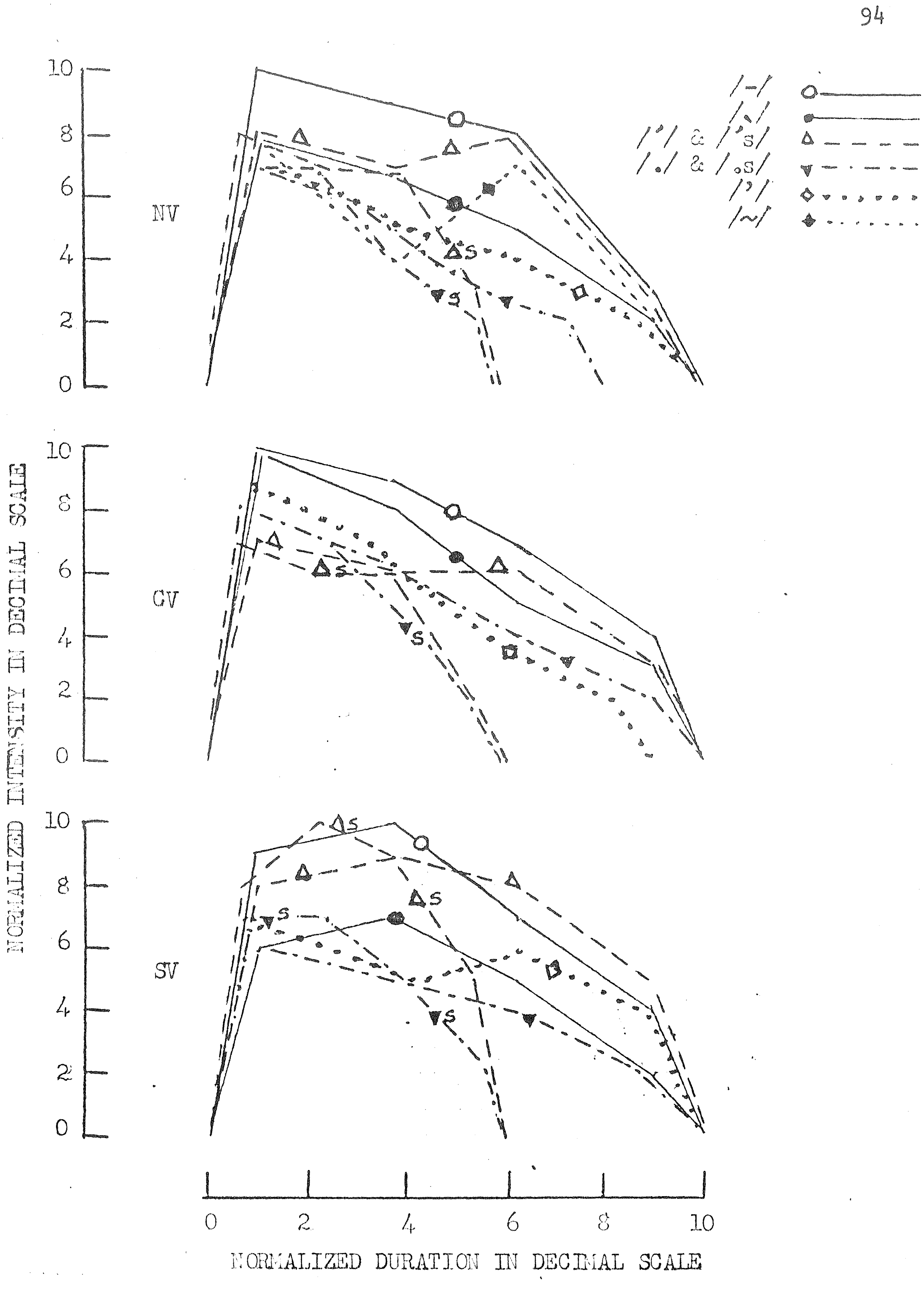


FIGURE 2.18 Normalized Mean Intensity of NV, CV and SV Tones Plotted Against Normalized Duration (Data from Table 2.15)

between various informants are summarized in Table 2.16 (next page) in which the Fo Differentials were calculated according to formula (N1), i.e. FD(0,E) stands for the FD between onset and endpoint, FD(0,M1) for the FD between onset and midpoint 1, etc.

Most of the variations occur as differences in the slopes of falling or rising contours, but in some cases they involve differences in contour as well. They may be characteristic of local subdialects or individual informants. In particular, the CV rising tone can be said to have two variants: one with a basically rising contour (8 informants) with only a slight fall at the beginning, not unlike the NV and SV rising tone; and one with a steeper fall and reaching fairly low Fo at midpoint 1. They may be local variants in many parts of the CV area.

2.4.3.3 Intensity The normalized values corroborate the earlier observations in 2.3.4 and permit us to compare the intensity curves of NV, CV and SV tones and describe them in unambiguous terms.

Let us take a difference of one decimal unit or less to mean 'sustained' and that of two decimal units or more to mean a fall or rise in intensity contour between the peak value (usually at I1 or I2) and I3 (I4 being always low in all tones), as presented in the Intensity columns of Table 2.15 (p 90). We can establish the intensity contours of all the tones as in Table 2.17, on page 97. Let us calculate the average intensity values in decimal units of each tone (the sum of I1, I2, I3, I4 divided by four), also from Table 2.15. The results suggest that three levels can be posited: high (average of 7.5 to 8.0 decimal units), mid (average of 6.2

TABLE 2.16 Examples of Variations in Fo Contour and Slope

Tone	Parameter	Popul. Means		Extreme Individual Means	
		Group	FD in %	Informant	FD in %
Level /-/	FD(0,E)	MV	-4	MM2	+6
				MM4	-10
		CV	+1	CM1	+20
		SV	-3	SM6	-10
Falling /ˎ/	FD(0,E)	MV	-12	MM1	-20
		CV	-10	CM1	-4
		SV	-10	SF1	-5
				SM5	-18
				SF3	-9 +5 (M2)*
Rising /ˊ/	FD(0,E)	MV	+30	MM6	+18
				MM1	+53
		CV	+14	CM8	+22
			-5 +20	CF4	-10 +46 (M1)*
		SV	+30	SM2	+15
				SF2	+53
Drop /./	FD(0,E)	MV	-14	MM2	-26
				MM3	-19 +12 (M2)*
		CV	-8	CM5	-3
				CM6	-15
				CM3	-12 +7 (M2)*
	FD(0,M1), FD(M1,E)	SV	-11 +12	SF2	-14 +27
				SM4	-4 (0,E)
Curve /ˀ/	FD(0,M2), FD(M2,E)	MV	-20 +18	MF2	-17 +8
				MM4	-23 +35
		CV	-11 +5	CM2	-1 +6
				CF4	-14 +23
				CM8	-25 (0,E)
	FD(0,M1), FD(M1,E)	SV	-14 +50	SM2	-11 +35
				SF3	-15 +84
Broken /~/	FD(0,M1), FD(M1,E)	MV	-16 +45	MM3	-21 +81
				MM4	-4 +27

* Concave contour with lowest Fo at Midpoint M1 or M2.

TABLE 2.17 Intensity Contour of NV, CV and SV Tones

Tone	NV	CV	SV
Level /-/	Falling (10-8)	Falling (10-7)	Falling (10-7) *
Falling /ˋ/	Falling (8-5)	Falling (10-5)	Falling (7-5)
Rising /ˊ/	Sustained (8-8)	Sustained (7-6)	Sustained (9-8)
St Ris /ˊs/	Sustained (8-7)	Sustained (7-6)	Sustained (10-9)
Drop /./	Falling (7-4)	Falling (8-4)	Falling (6-4)
St Drop /.s/	Falling (7-5)	Falling (8-5)	Falling (7-5)
Curve /’/	Falling (7-4)	Falling (9-4)	Sustained (7-6)
Broken /~/	Concave (8-4-7)		

* Figures in parentheses indicate changes in normalized values of intensity (decimal units) between peak intensity and I3 (about 2/3 of duration when I starts to decay in all tones).

TABLE 2.18 Average Intensity Level of NV, CV and SV Tones

Tone	NV	CV	SV
Level /-/	High (7.5)	High (7.5)	High (7.5) *
Falling /ˋ/	Low (5.5)	Mid (6.5)	Low (5.0)
Rising /ˊ/	Mid (6.5)	Low (5.5)	High (7.5)
St Ris /ˊs/	Mid (6.2)	Low (5.2)	High (8.0)
Drop /./	Low (4.7)	Low (5.0)	Low (4.2)
St Drop /.s/	Low (5.2)	Low (5.5)	Low (5.2)
Curve /’/	Low (4.5)	Low (5.5)	Low (5.5)
Broken /~/	Low (5.2)		

* Figures in parentheses indicate average intensity levels in normalized values : high (7.5-8.0); mid (6.2-6.5); low (4.2-5.5).

to 6.5 decimal units) and low (average of 4.2 to 5.5 decimal units). Table 2.18 (p 97) gives the intensity levels of all tones.

The following observations can be made from Tables 2.17 and 2.18:

(a) There is great similarity in intensity contour between the same tones across dialects, the only exception being the SV curve tone (see Table 2.17). There is less similarity in intensity levels (only four out of seven common tones have the same levels; see Table 2.18) This strengthens the idea that tone contour, primarily manifested in Fo and secondarily in intensity, is the dominant and unifying feature in the Vietnamese tone system across dialects.

(b) The normalized values and comparisons in the abovementioned tables corroborate the notion of partial correlation between intensity and Fo, as noted in 2.3.4, especially during the first two thirds of duration. Level, falling and low concave tones tend to have falling intensity contours. Rising tones , and the SV mid-concave tone with high rising end, tend to have sustained intensity contours. Only the NV broken tone has a clear concave intensity contour. This suggests that intensity is not an independent variable, and with the exception of the NV broken tone, is not likely to play an independent role in tone recognition.

2.4.3.4 <u>Duration</u> Normalized values for duration establish a class of short tones in all dialects: the stopped tones. These have a normalized duration of 6 decimal units, while the long, non-stopped tones have the normalized duration of 10. The two creaky-ending tones, the NV drop tone and the CV curve tone, have normalized duration of 8 and 9 respectively. This regularity suggests that duration is not an independent variable in Vietnamese tones, because it is always associated with the presence or absence of voiceless final stops or creaky endings, therefore fully predictable regarding its relative value in similar contexts. Duration and intensity are therefore excluded from phonological feature specification of the tones on similar grounds.

2.5 CONCLUSION

In this chapter, I have presented the results of an acoustically based study of Vietnamese tones in the three major dialects and have indicated some possible subdialectal and individual variations. Normalization procedures have also been proposed and applied for the comparison of data. They will be useful for further research into more variations in local subdialects. Four sets of parameters, Fo, laryngealization, intensity and duration, are used for describing tonal characteristics in their varying forms across dialects. They are flexible and comprehensive enough, I feel, to be used in describing the wider variations of tone in connected speech. The data establish four types of contour (level, falling, rising and concave), three levels of average Fo (high, mid and low), and four degrees of laryngealization as features characterizing Vietnamese tones at the physical phonetic level in all dialects.

The tone systems of the three major dialects display both similarities and differences in terms of the above-mentioned parameters and features. The most important similarity seems to be Fo contour, since all but one of the tones have basically similar contours across dialects, while varying in relative Fo level or laryngealization. The remaining one, the drop tone, displays difference in Fo contour in one dialect only, SV, as laryngealization marks its variation between the other two dialects.

The above observations, together with the fact that Vietnamese has only one phonetic level tone, suggest the following conclusions:

(a) Fo contour is the most important and the most basic feature in the Vietnamese tone system. This implies that it may be the most important cue for tone recognition within a dialect and across dialects. This point will be dealt with in more detail in Ch. Four.

(b) Vietnamese can thus be characterized as belonging to a type of "contour tone language with register overlap" as defined by Pike (148:12-13). More precisely, on the basis of the essential phonetic properties described in this chapter, systematic phonetic and phonological descriptions of Vietnamese tones should include features of pitch contour, relative pitch level and laryngealization (or voice quality).

Furthermore, the variations of Fo and duration in different environments are interesting aspects which deserve further examination if we are to understand the relationships between them and the nature of tone production in Vietnamese.

APPENDIX TO CHAPTER TWO

Details of Fo Data and

Illustrative Mingograms and Spectrograms

of NV, CV and SV Tones

by Individual Informants

TABLE A.1 Mean Fo in Hertz and Standard Deviations of NV Level Tone /-/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
NF1	8	245	246	243	240	233	226
		12	12	6	9	10	8*
NF2	8	292	290	289	286	283	280
		7	7	7	8	6	5
NF3	8	251	257	266	262	254	246
		17	15	11	18	20	21
NF4	8	239	242	241	238	235	232
		20	17	13	16	20	18
NM1	8	138	136	139	137	130	127
		12	10	9	7	7	5
NM2	8	156	157	160	161	162	166
		22	18	17	19	22	21
NM3	8	243	243	244	241	239	229
		13	4	4	5	7	9
NM4	8	134	129	129	125	129	121
		13	3	2	0	7	4
NM6	8	214	208	205	203	198	197
		13	6	7	4	6	9

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.2 Mean Fo in Hertz and Standard Deviations of NV Falling Tone /˷/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
NF1	8	211	207	206	197	194	192
		7	4	6	8	6	6*
NF2	8	229	227	218	212	211	209
		13	11	12	12	8	9
NF3	8	226	226	220	209	205	198
		15	11	7	7	4	8
NF4	8	211	209	201	192	186	178
		32	29	20	12	9	10
NM1	8	125	121	116	108	104	99
		6	6	6	4	5	7
NM2	8	124	121	122	115	107	106
		10	9	9	12	18	15
NM3	8	186	186	183	178	174	167
		8	8	8	6	9	7
NM4	8	114	106	106	101	96	92
		6	3	3	3	3	4
NM6	8	178	174	170	163	158	157
		10	7	5	5	5	6

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.3 Mean Fo in Hertz and Standard Deviations of NV Rising Tone /ʹ/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
NF1	8	221	221	229	259	298	298
		9	6	5	12	9	11*
NF2	8	275	267	269	286	307	321
		19	15	8	11	27	35
NF3	8	236	234	240	261	307	314
		18	17	17	17	25	22
NF4	8	232	226	228	246	269	296
		19	15	13	15	22	40
NM1	8	130	123	124	149	179	199
		13	7	7	3	15	27
NM2	8	143	139	141	159	191	217
		21	20	17	15	13	17
NM3	8	227	222	221	241	264	274
		17	10	6	9	19	29
NM4	8	127	124	127	149	177	171
		16	6	6	11	8	16
NM6	8	195	192	196	209	221	231
		12	6	5	8	13	11

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.4 Mean Fo in Hertz and Standard Deviations of NV Stopped Rising Tone /'s/ at Four Timepoints

Informant	n	P1	P2	P3	P4
NF1	8	259	261	275	287
		19	12	7	6*
NF2	8	318	327	340	354
		17	15	12	9
NF3	8	240	257	287	329
		27	19	12	7
NF4	8	251	265	290	314
		25	21	17	20
NM1	8	137	149	165	182
		11	9	11	17
NM2	8	157	172	191	211
		22	13	10	19
NM3	8	258	269	281	291
		22	12	6	7
NM4	8	147	155	177	184
		16	11	12	13
NM6	8	225	236	248	261
		19	11	6	3

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.5 Mean Fo in Hertz and Standard Deviations of NV Drop Tone /./ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
NF1	8	219	218	217	215	206	202
		19	12	9	8	4	4*
NF2	8	246	243	229	211	202	208
		14	12	14	15	26	24
NF3	8	236	234	230	214	213	214
		20	14	12	8	9	11
NF4	8	217	217	211	181	181	187
		23	20	17	21	13	27
NM1	8	129	125	121	111	107	108
		13	11	8	7	9	11
NM2	8	136	129	125	114	106	100
		16	11	10	9	8	10
NM3	8	201	197	188	162	171	181
		14	9	7	11	16	17
NM4	8	126	121	118	111	100	100
		12	6	8	13	8	16
NM6	8	191	193	189	175	170	168
		11	5	4	11	14	14

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.6 Mean Fo in Hertz and Standard Deviations of NV Stopped Drop Tone /.s/ at Four Timepoints

Informant	n	P1	P2	P3	P4
NF1	8	216	209	201	197
		8	3	5	4*
NF2	8	234	216	206	199
		18	20	13	11
NF3	8	224	221	212	205
		12	8	7	11
NF4	8	203	200	189	176
		12	10	11	10
NM1	8	123	118	113	101
		5	4	4	3
NM2	8	131	124	119	109
		16	11	11	8
NM3	8	200	192	187	179
		15	12	7	11
NM4	8	116	111	107	94
		10	7	7	8
NM6	8	192	182	176	166
		14	8	7	12

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.7 Mean Fo in Hertz and Standard Deviations of NV Curve Tone /'/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
NF1	8	215	214	201	179	187	207
		10	9	7	6	7	17 *
NF2	8	226	223	207	187	193	202
		21	16	11	10	14	19
NF3	8	225	217	199	176	214	219
		17	13	8	2	12	11
NF4	8	204	197	182	157	156	162
		29	26	21	21	14	18
NM1	8	124	119	108	98	118	134
		5	5	6	4	11	20
NM2	8	121	117	106	91	101	114
		9	6	5	6	4	12
NM3	8	184	178	163	146	157	166
		22	17	9	4	11	14
NM4	8	117	105	97	90	114	122
		13	5	5	5	16	20
NM6	8	174	167	152	144	158	169
		12	9	5	7	7	8

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.8 Mean Fo in Hertz and Standard Deviations of NV Broken Tone /~/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
NF1	8	232	227	195	251	284	287
		14	10	18	14	10	23*
NF2	8	286	277	248	282	315	322
		31	31	43	33	18	22
NF3	8	237	232	184	253	303	307
		20	21	20	9	10	9
NF4	8	233	227	186	240	265	267
		18	21	16	9	16	18
NM1	8	132	130	107	134	169	176
		12	11	18	13	4	11
NM2	8	139	133	110	152	187	199
		14	13	19	7	8	9
NM3	8	227	226	192	252	280	272
		21	14	7	18	23	20
NM4	8	132	129	127	146	178	162
		20	13	6	11	17	13
NM6	8	199	183	173	192	214	212
		22	17	13	16	13	23

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.9 Mean Fo in Hertz and Standard Deviations of CV Level Tone /-/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
CF1	8	267	254	254	259	262	254
		18	11	10	11	11	7*
CF2	8	267	271	276	277	276	274
		9	8	6	8	13	16
CF3	8	254	248	249	251	250	251
		13	7	7	9	7	10
CF4	8	219	212	216	226	231	227
		19	12	12	8	9	7
CM1	8	175	175	186	193	203	211
		11	8	5	5	11	29
CM2	8	142	142	146	147	148	145
		13	11	11	8	7	9
CM3	8	124	121	127	139	137	130
		10	6	3	8	5	8
CM4	8	146	150	154	157	157	157
		4	0	3	5	6	8
CM5	8	138	136	141	137	132	131
		6	4	6	4	4	6
CM6	8	129	128	127	129	129	129
		6	5	4	4	6	5
CM7	8	169	168	170	166	162	165
		7	6	6	5	4	5
CM8	8	178	179	179	177	173	169
		10	9	10	11	13	15

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.10 Mean Fo in Hertz and Standard Deviations of CV Falling Tone /˴/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
CF1	8	251	241	237	234	229	222
		18	17	15	12	9	12*
CF2	8	252	252	249	242	234	227
		11	8	6	7	6	5
CF3	8	239	235	224	219	217	214
		11	10	13	12	11	12
CF4	8	202	199	194	189	184	182
		7	4	5	5	6	7
CM1	8	181	182	179	172	170	172
		7	5	4	4	3	12
CM2	8	124	123	120	116	113	111
		6	6	5	6	6	4
CM3	8	124	121	122	121	120	115
		5	4	4	2	4	6
CM4	8	144	142	139	136	134	130
		5	4	3	4	3	6
CM5	8	131	128	122	117	116	116
		11	10	6	6	7	7
CM6	8	117	114	112	107	106	101
		7	5	4	4	3	3
CM7	8	158	156	152	147	144	144
		7	6	5	4	4	4
CM8	8	167	166	161	154	147	142
		11	10	9	9	9	10

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.11 Mean Fo in Hertz and Standard Deviations of CV Rising Tone /ʹ/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
CF1	8	210	199	198	204	214	212
		9	7	7	8	16	29 *
CF2	8	218	213	211	222	248	263
		8	6	6	11	14	18
CF3	8	239	226	209	221	247	259
		12	11	12	13	7	9
CF4	8	177	169	159	176	215	232
		12	9	6	9	25	34
CM1	8	157	151	122	143	163	163
		22	19	21	13	12	15
CM2	8	124	121	124	134	142	146
		7	3	4	6	5	5
CM3	8	117	112	102	117	119	119
		7	4	4	5	2	13
CM4	8	124	122	127	134	137	137
		8	6	4	6	4	5
CM5	8	121	121	124	136	143	146
		6	5	9	13	11	9
CM6	8	113	114	117	123	127	128
		7	6	6	7	7	7
CM7	8	167	161	168	190	212	212
		12	6	6	12	23	17
CM8	8	131	133	142	150	157	160
		8	7	6	7	10	12

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.12 Mean Fo in Hertz and Standard Deviations of CV Stopped Rising Tone /ʹs/ at Four Timepoints

Informant	n	P1	P2	P3	P4
CF1	8	211	202	200	204
		12	7	6	9*
CF2	8	222	221	237	264
		9	7	12	23
CF3	8	234	227	235	253
		14	10	7	9
CF4	8	188	166	169	208
		14	15	10	11
CM1	8	162	141	151	168
		8	16	12	14
CM2	8	132	138	147	157
		12	11	7	6
CM3	8	112	113	121	116
		7	5	3	6
CM4	8	136	132	136	140
		10	6	6	5
CM5	8	127	124	127	132
		10	6	2	4
CM6	8	114	119	129	140
		5	5	8	15
CM7	8	176	178	188	197
		11	7	8	7
CM8	8	143	149	161	171
		13	10	12	17

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.13 Mean Fo in Hertz and Standard Deviations of CV Drop Tone /./ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
CF1	8	222	208	202	199	199	203
		12	6	6	6	5	15*
CF2	8	225	221	218	212	207	206
		10	7	3	3	2	3
CF3	8	228	221	214	210	211	212
		7	8	4	5	6	4
CF4	8	188	181	175	171	166	172
		12	9	7	5	5	6
CM1	8	171	167	157	149	156	159
		10	6	7	3	15	22
CM2	8	122	119	114	107	104	104
		7	5	4	4	4	3
CM3	8	119	116	116	104	108	111
		7	4	5	8	5	4
CM4	8	137	132	126	122	119	119
		12	10	7	5	4	6
CM5	8	120	117	116	114	114	116
		3	5	2	2	5	4
CM6	8	120	116	113	109	106	102
		10	6	4	4	3	4
CM7	8	154	149	142	140	149	154
		5	2	4	9	11	12
CM8	8	157	156	151	141	139	138
		10	8	8	8	10	12

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.14 Mean Fo in Hertz and Standard Deviations of CV Stopped Drop Tone /.s/ at Four Timepoints

Informant	n	P1	P2	P3	P4
CF1	8	226	208	204	197
		13	5	7	5*
CF2	8	239	231	224	220
		8	7	6	7
CF3	8	232	221	216	212
		8	6	5	5
CF4	8	192	178	171	166
		10	2	3	4
CM1	8	176	166	156	148
		7	6	7	4
CM2	8	124	121	118	112
		7	3	2	7
CM3	8	125	122	118	102
		7	2	4	4
CM4	8	142	127	131	127
		6	2	2	5
CM5	8	131	124	120	119
		9	7	4	3
CM6	8	121	114	108	103
		6	3	4	7
CM7	8	159	151	144	151
		8	9	6	9
CM8	8	168	161	151	146
		12	6	5	6

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.15 Mean Fo in Hertz and Standard Deviations of CV Curve Tone /ʼ/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
CF1	8	219	213	201	193	196	199
		12	9	6	7	7	9 *
CF2	8	237	232	224	221	217	216
		12	6	7	10	18	26
CF3	8	240	236	226	223	231	236
		10	9	7	12	28	36
CF4	8	193	187	177	166	197	204
		9	6	4	6	15	18
CM1	8	170	169	159	142	154	168
		7	4	4	10	23	24
CM2	8	123	121	121	121	125	129
		10	8	4	6	11	15
CM3	8	117	112	110	95	99	101
		9	5	5	5	6	8
CM4	8	141	139	132	122	126	129
		5	2	4	7	7	7
CM5	8	124	123	121	115	117	119
		7	5	4	5	3	3
CM6	8	116	113	109	104	101	101
		9	7	7	5	3	5
CM7	8	159	154	147	139	147	156
		7	6	5	3	5	5
CM8	8	165	159	147	137	126	124
		10	11	10	12	14	16

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.16 Mean Fo in Hertz and Standard Deviations of SV Level Tone /-/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
SF1	8	294	294	293	292	294	287
		4	4	4	5	5	11 *
SF2	8	209	207	211	216	211	200
		27	21	12	13	21	18
SF3	8	231	234	237	236	229	226
		18	14	15	16	18	23
SF5	8	261	258	257	254	254	256
		12	10	13	10	11	12
SM1	8	205	211	209	208	206	198
		11	6	6	7	9	16
SM2	8	136	135	135	134	132	132
		7	6	7	5	6	3
SM4	8	124	124	125	126	128	125
		4	4	5	7	8	6
SM5	8	142	142	142	141	140	136
		10	3	4	6	6	7
SM6	8	124	122	121	119	116	111
		6	5	4	5	8	10

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.17 Mean Fo in Hertz and Standard Deviations of SV Falling Tone /\/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
SF1	8	271	269	263	259	258	256
		13	10	8	5	6	7*
SF2	8	172	161	152	147	141	139
		25	12	7	5	4	6
SF3	8	196	190	185	179	185	188
		10	7	7	9	13	8
SF5	8	213	204	199	194	191	204
		16	8	5	6	7	12
SM1	8	171	170	166	161	157	156
		9	9	7	4	5	4
SM2	8	127	119	114	108	106	104
		7	3	3	2	3	2
SM4	8	109	107	105	103	102	99
		6	5	5	4	5	3
SM5	8	132	127	121	116	111	107
		7	6	5	4	3	6
SM6	8	106	103	99	96	93	92
		6	4	4	4	4	2

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.18 Mean Fo in Hertz and Standard Deviations of SV Rising Tone /ʹ/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
SF1	8	304	296	302	326	353	364
		20	13	9	11	15	19*
SF2	8	187	184	201	248	299	287
		28	19	25	29	29	34
SF3	8	235	233	251	286	325	336
		19	15	16	22	11	34
SF5	8	263	259	266	296	349	366
		13	12	9	14	23	32
SM1	8	214	216	226	242	256	247
		11	9	9	9	10	20
SM2	8	142	142	146	156	164	164
		4	4	4	6	7	11
SM4	8	120	121	128	144	154	168
		7	4	10	4	4	7
SM5	8	153	149	159	175	189	204
		21	15	17	17	10	18
SM6	8	129	130	136	149	162	159
		11	8	6	6	10	9

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.19 Mean Fo in Hertz and Standard Deviations of SV Stopped Rising Tone /ˊs/ at Four Timepoints

Informant	n	P1	P2	P3	P4
SF1	8	328	327	356	391
		22	9	13	14*
SF2	8	194	215	244	297
		25	10	13	13
SF3	8	239	251	277	321
		24	21	18	18
SF5	8	267	271	295	363
		10	7	7	14
SM1	8	226	234	244	257
		11	7	10	15
SM2	8	145	146	154	164
		6	4	5	5
SM4	8	127	131	151	169
		5	5	2	8
SM5	8	160	167	189	191
		21	17	18	8
SM6	8	130	132	143	157
		15	11	13	10

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.20 Mean Fo in Hertz and Standard Deviations of SV Drop Tone /./ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
SF1	8	281	264	252	259	269	268
		9	5	12	6	8	9*
SF2	8	154	142	133	147	170	169
		15	7	6	4	5	3
SF3	8	199	189	174	182	196	204
		8	7	4	4	6	11
SF5	8	220	202	194	200	224	235
		16	7	6	9	15	12
SM1	8	184	171	156	144	156	163
		14	14	11	4	6	10
SM2	8	120	114	109	108	116	121
		7	5	2	6	8	11
SM4	8	108	107	104	103	106	104
		6	6	5	3	3	3
SM5	8	132	125	118	116	124	127
		8	6	6	5	5	4
SM6	8	102	100	97	96	101	108
		5	3	4	3	3	6

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.21 Mean Fo in Hertz and Standard Deviations of SV Stopped Drop Tone /.s/ at Four Timepoints

Informant	n	P1	P2	P3	P4
SF1	8	277	264	256	254
		16	7	5	7*
SF2	8	165	147	152	167
		22	5	5	12
SF3	8	207	192	185	196
		18	6	6	8
SF5	8	221	206	199	221
		13	7	5	12
SM1	8	187	180	169	166
		14	14	16	11
SM2	8	127	119	116	111
		7	3	4	2
SM4	8	112	106	106	111
		8	5	4	4
SM5	8	132	124	120	121
		9	4	5	7
SM6	8	104	102	99	99
		9	6	2	6

* For each informant, mean Fo values on first line, SD on second line.

TABLE A.22 Mean Fo in Hertz and Standard Deviations of SV Curve Tone /ʾ/ at Six Timepoints

Informant	n	P1	P2	P3	P4	P5	P6
SF1	8	307	287	256	276	315	317
		14	26	7	7	12	14 *
SF2	8	156	146	137	165	241	246
		12	12	7	17	16	22
SF3	8	202	187	172	223	294	317
		14	9	6	5	11	27
SF5	8	231	211	195	216	299	310
		19	17	5	6	16	24
SM1	8	189	177	152	191	214	210
		4	7	4	10	14	20
SM2	8	125	115	111	127	149	150
		7	5	3	8	6	5
SM4	8	111	104	100	122	149	156
		8	4	5	5	6	7
SM5	8	137	128	122	142	178	167
		16	10	7	11	18	16
SM6	8	106	101	96	112	136	144
		9	4	4	6	7	11

* For each informant, mean Fo values on first line, SD on second line.

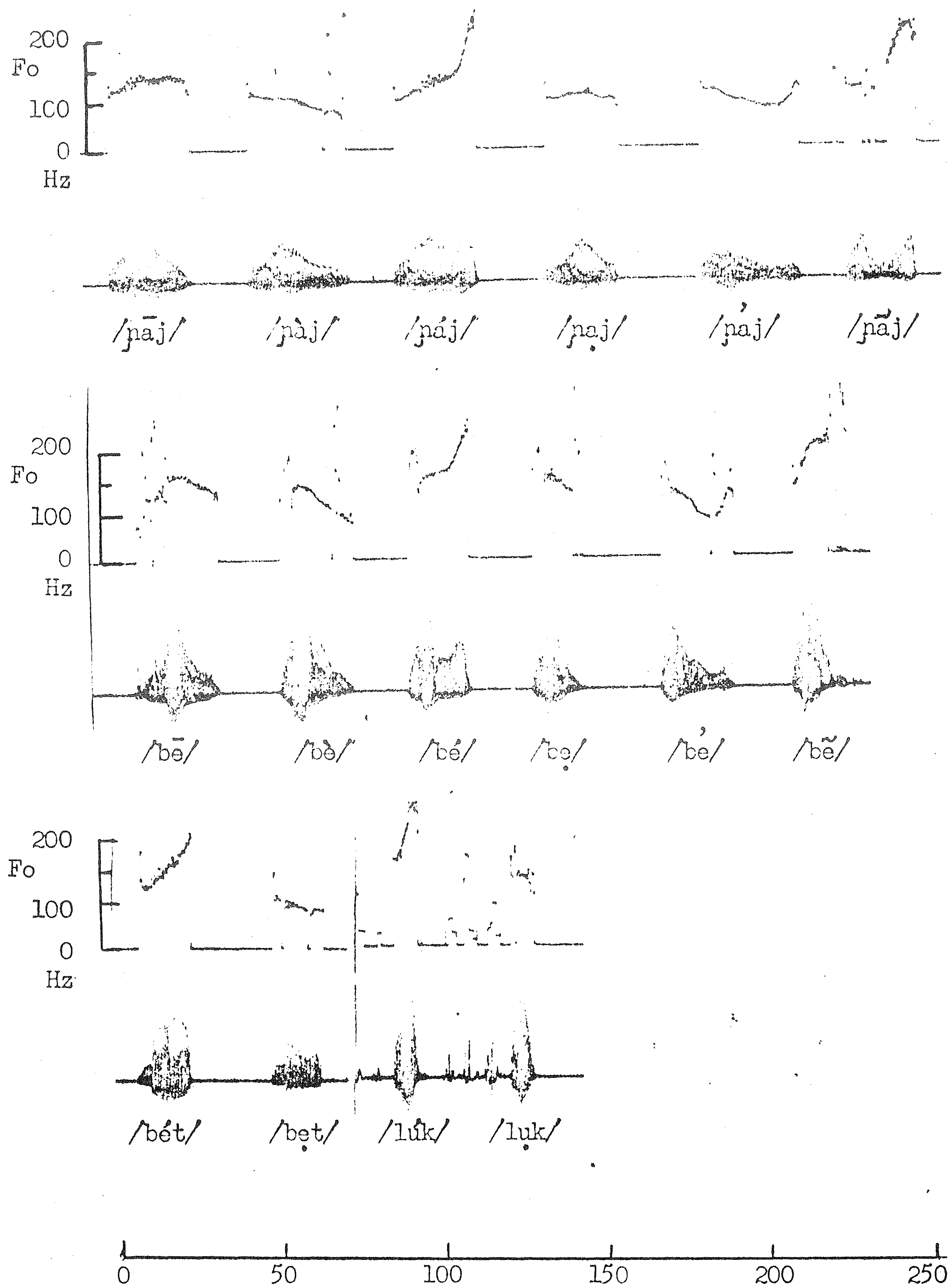


FIGURE A.1 Fo Curves and Waveforms of Different Syllables with NV Tones by Informant NM4 Showing Basic Tone Contours and Variations in Fo Slopes and Waveform Amplitudes

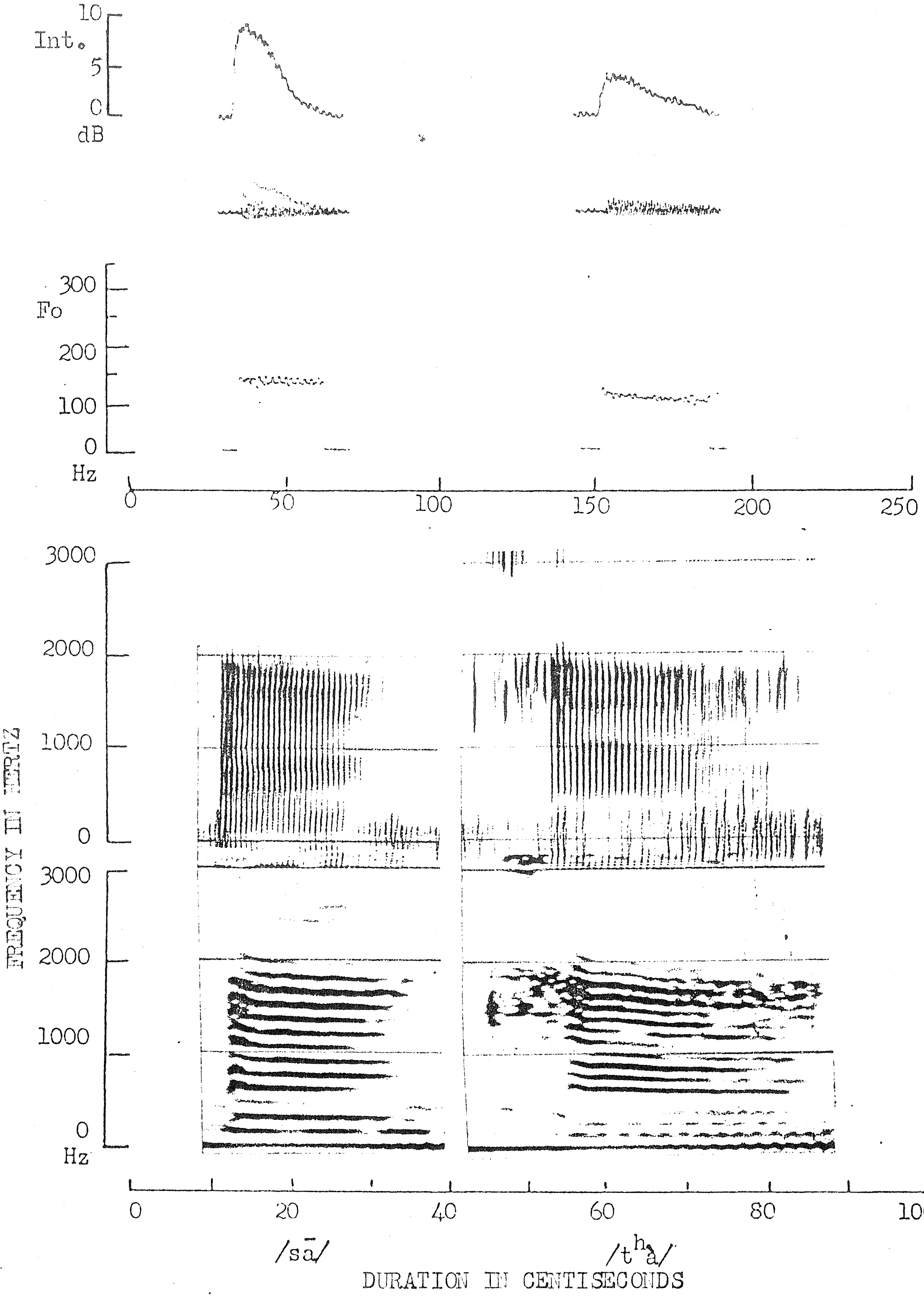


FIGURE A.2 Mingograms and Spectrograms of Syllables with CV Level and Falling Tones by Informant CM2

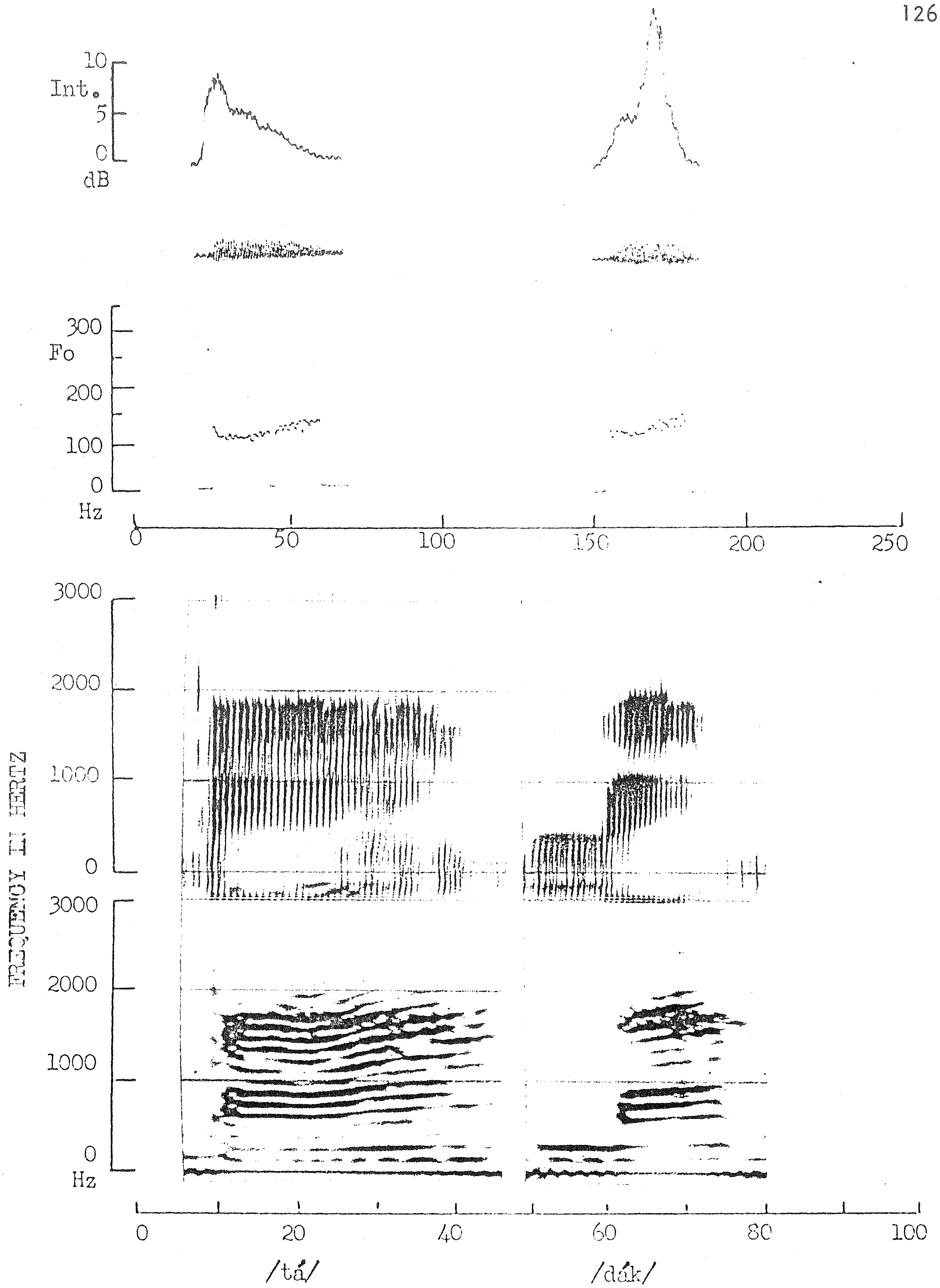


FIGURE A.3 Mingograms and Spectrograms of Syllables with CV Rising and Stopped Rising Tones by Informant CM2

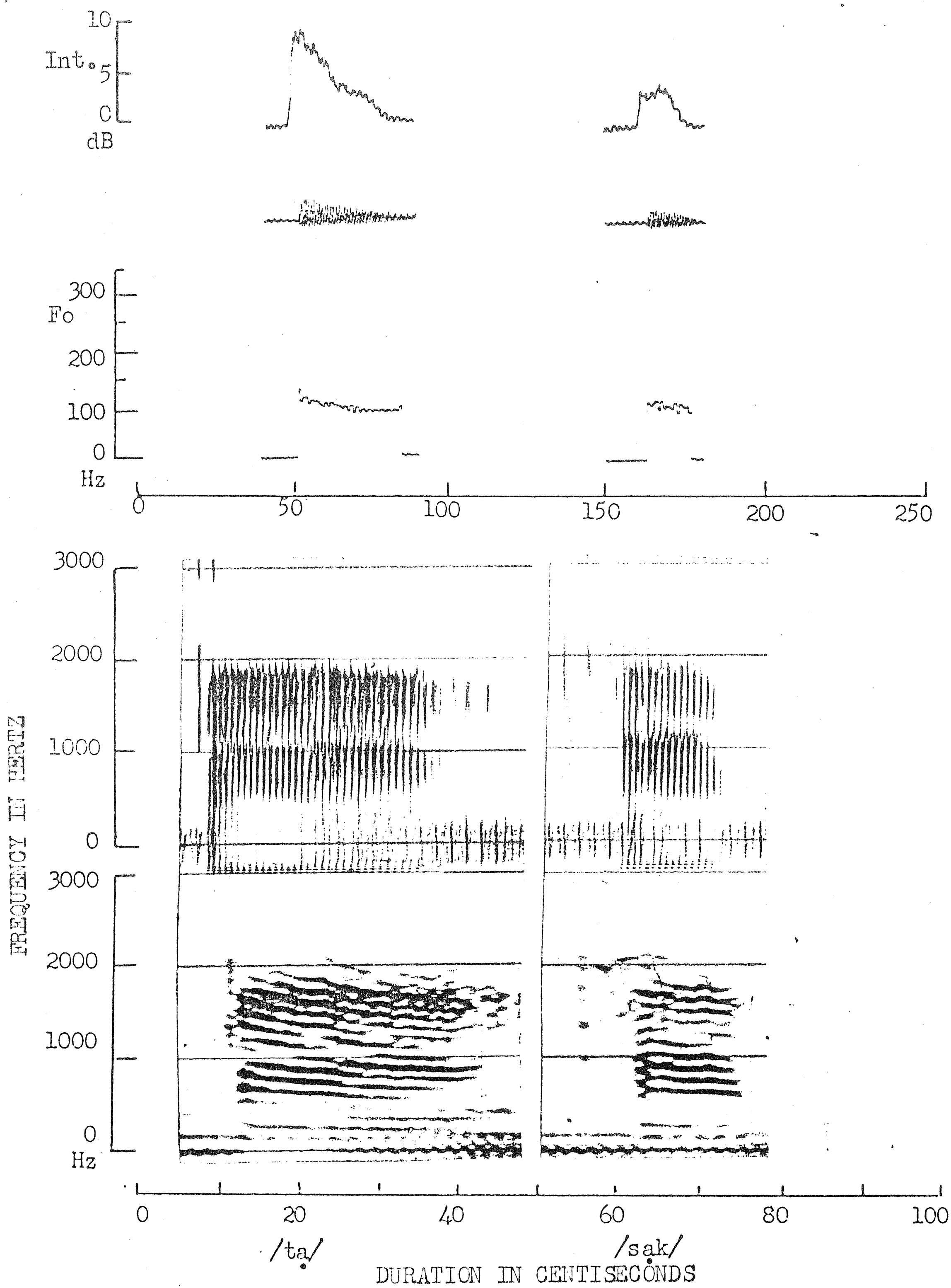


FIGURE A.4 Mingograms and Spectrograms of Syllables with CV Drop and Stopped Drop Tones by Informant CM2

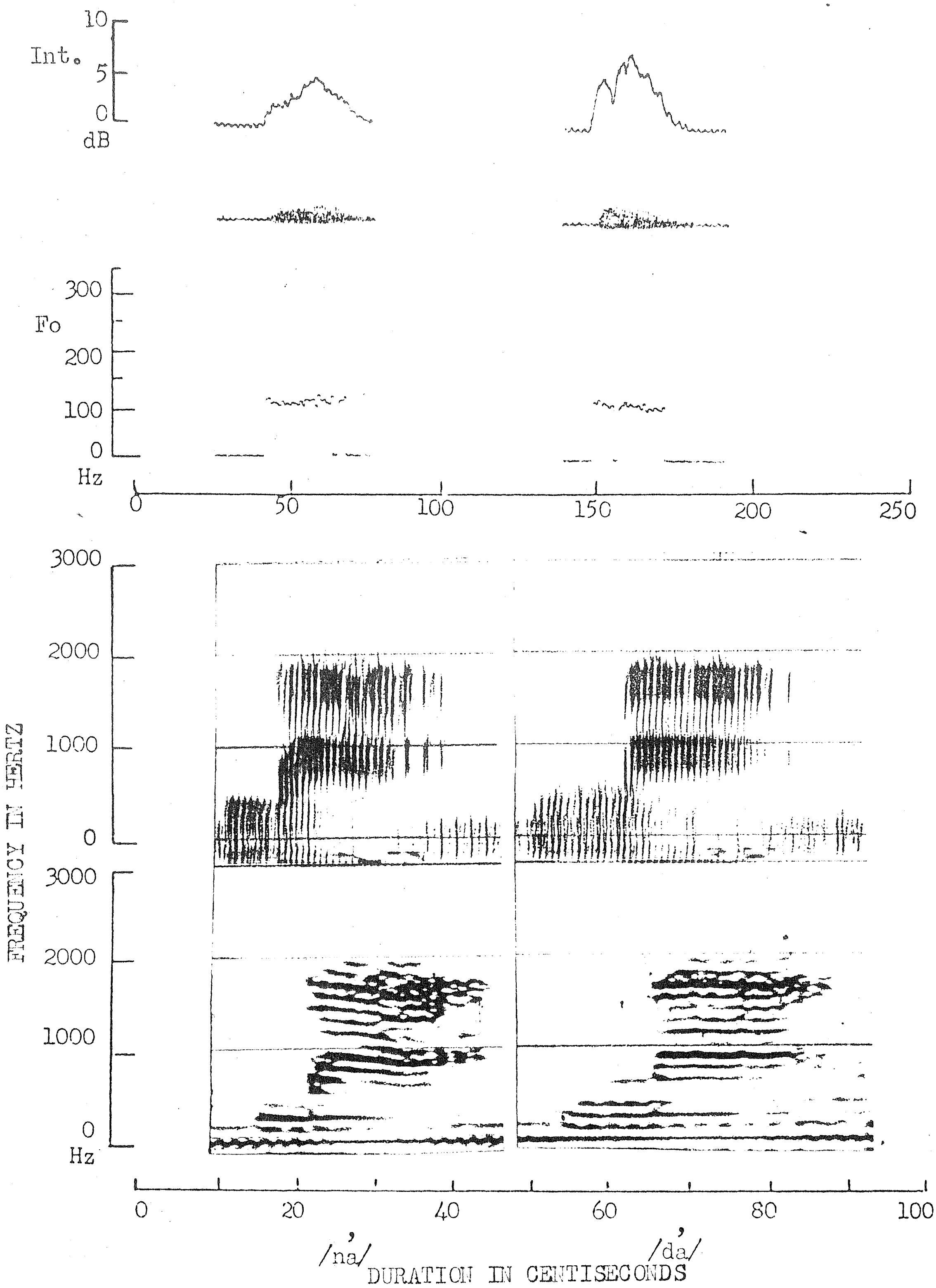


FIGURE A.5 Mingograms and Spectrograms of Syllables with CV Curve Tone by Informant CM2

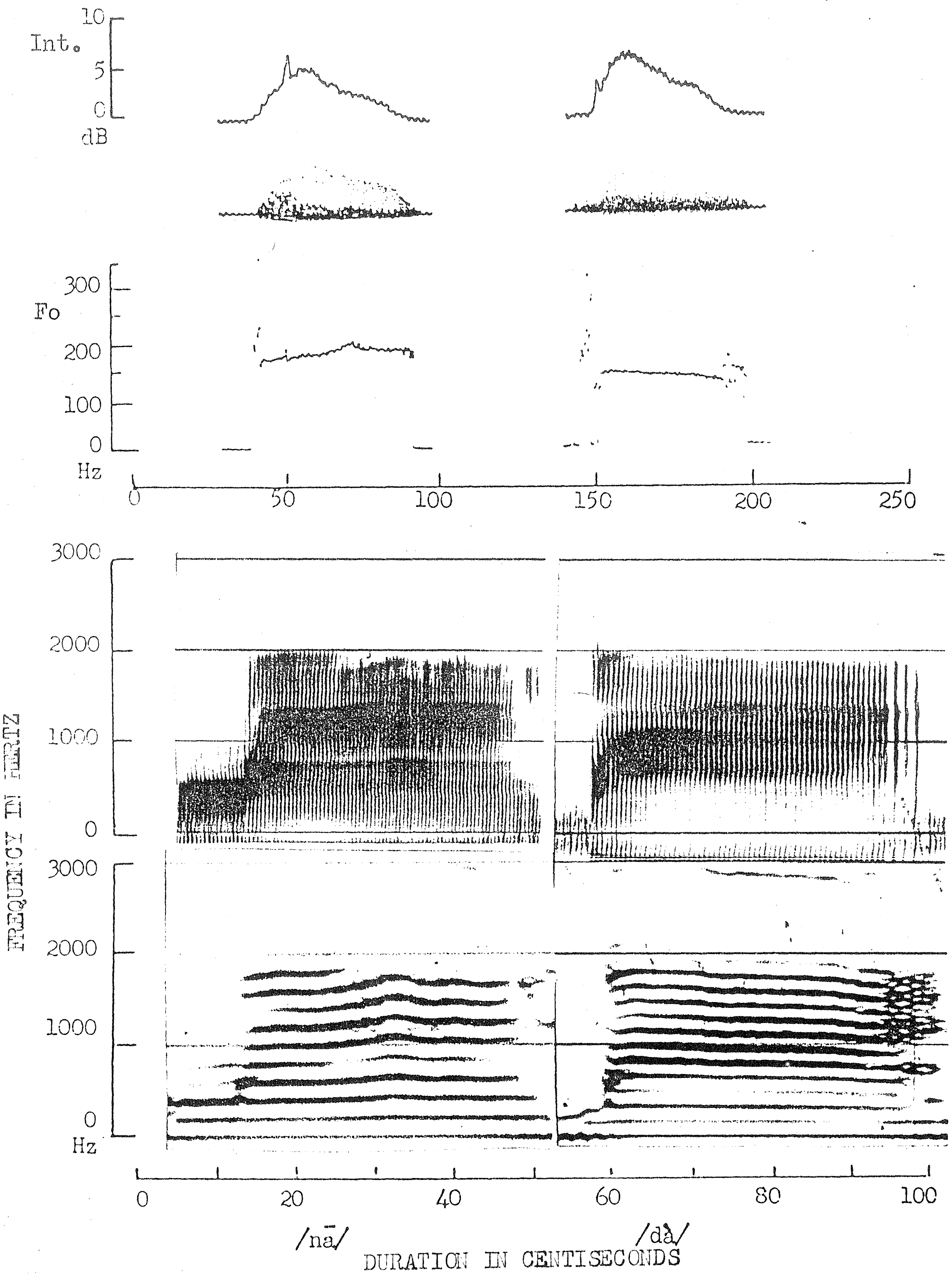


FIGURE A.6 Mingograms and Spectrograms of Syllables with SV Level and Falling Tones by Informant SF2

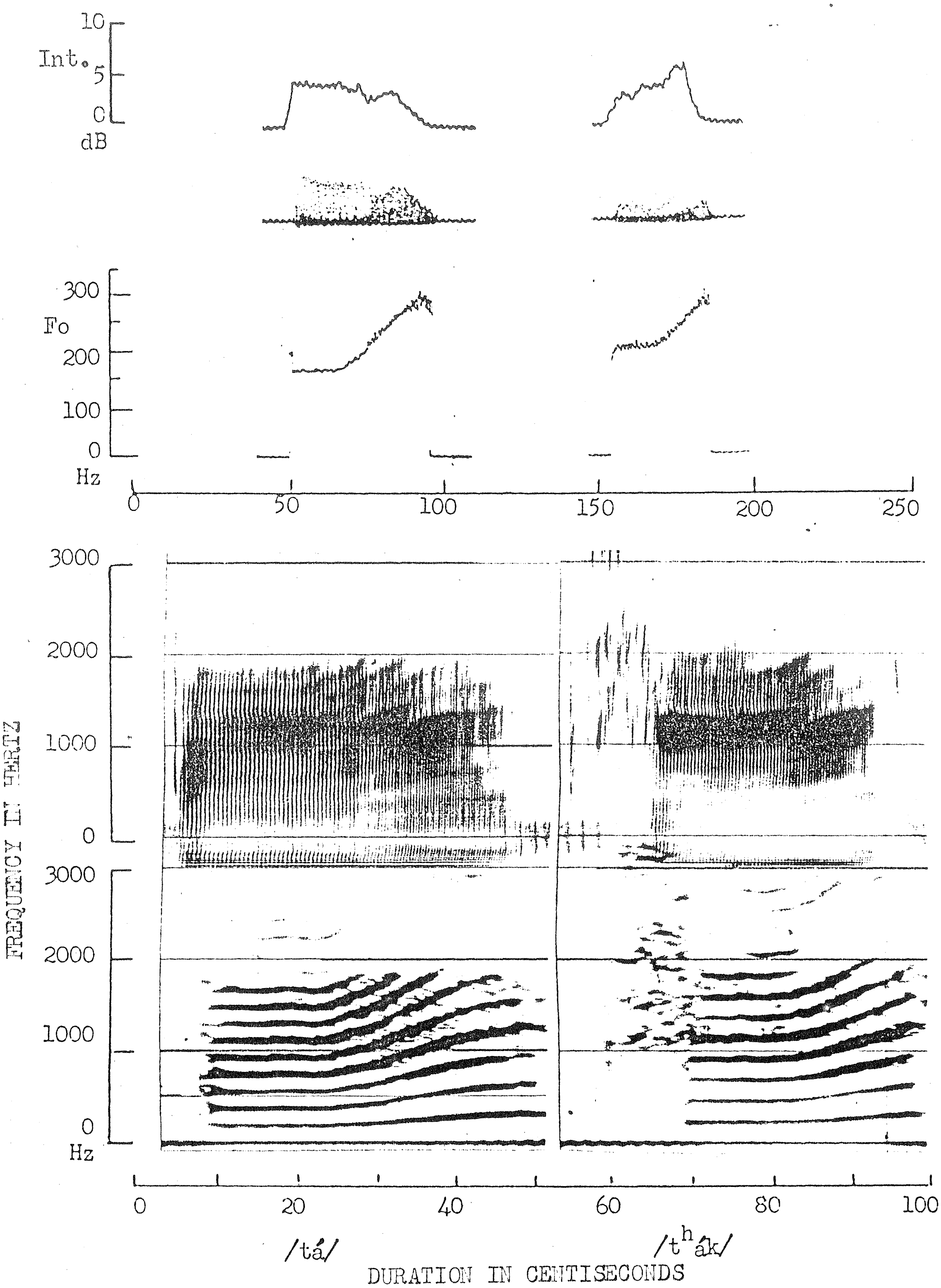


FIGURE A.7 Mingograms and Spectrograms of Syllables with SV Rising and Stopped Rising Tones by Informant SF2

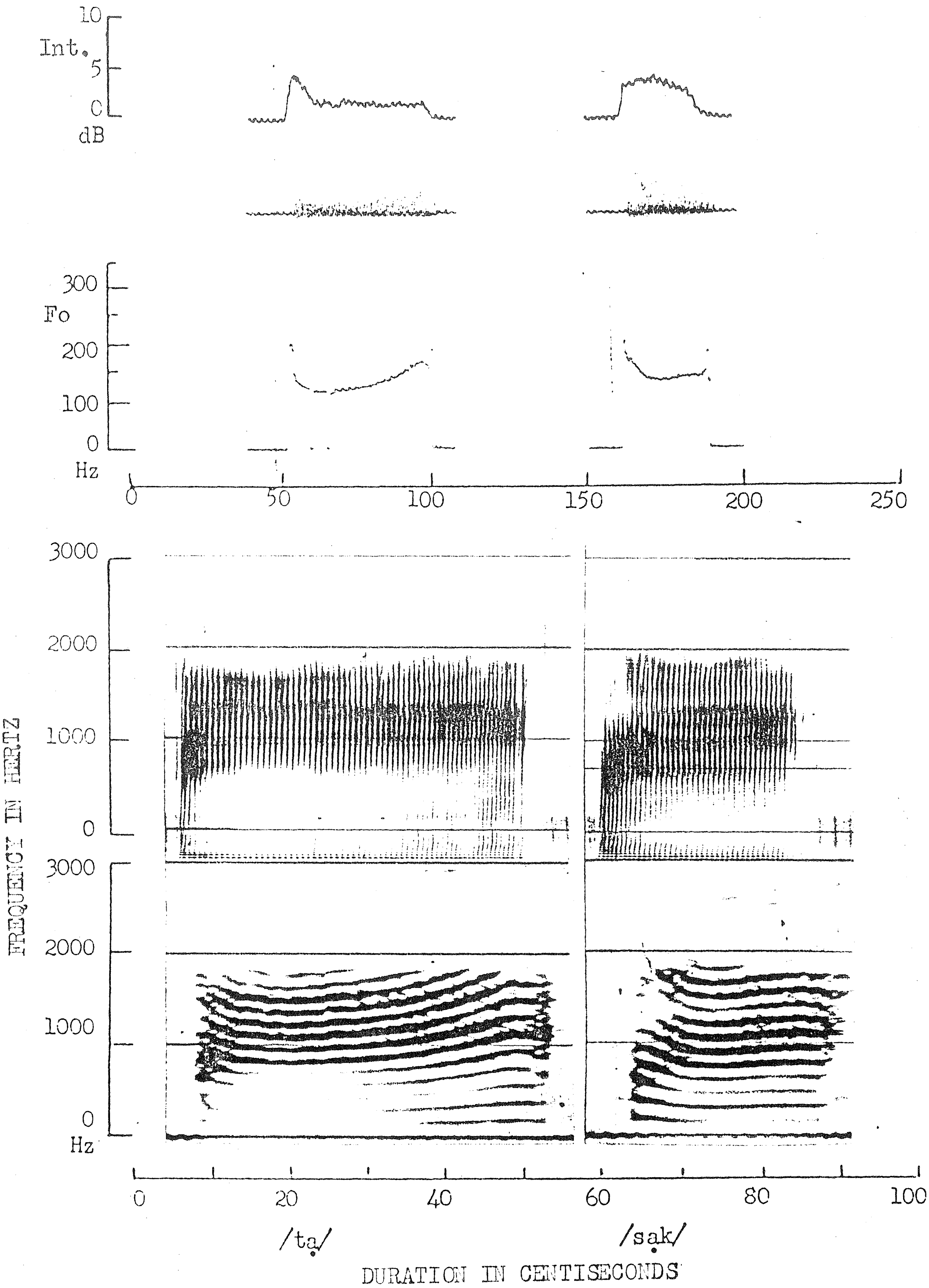


FIGURE A.8 Mingograms and Spectrograms of Syllables with SV Drop and Stopped Drop Tones by Informant SF2

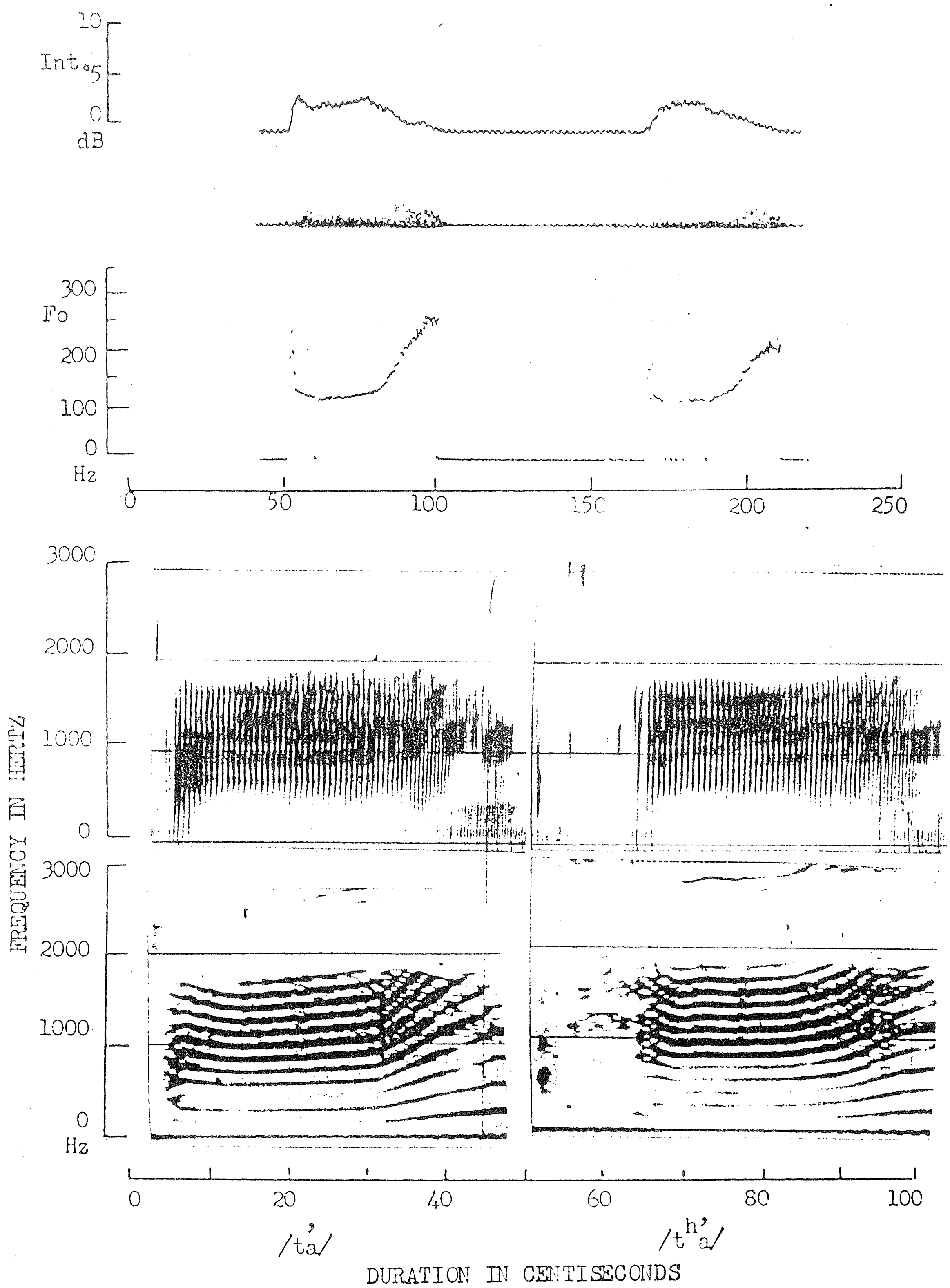


FIGURE A.9 Mingograms and Spectrograms of Syllables with SV Curve Tone by Informant SF2

CHAPTER THREE

Fo VARIATION AND THE PHONATION PROCESS

The description of acoustic phonetic properties of Vietnamese tones in Ch. Two has shown that in spite of their primary importance, Fo parameters alone cannot properly characterize all tonal differences in the Vietnamese system. Subdialectal and individual variations further suggest that the relationships between Fo, intensity, duration and laryngealization may be quite complex and further investigations are needed to understand their nature. In this chapter, I shall examine some more data concerning aspects of those relationships and offer some speculations on the possible causes of the observed tonal and related phenomena in terms of their physiological and physical nature.

3.1. *SOME MEANINGFUL RELATIONSHIPS*

3.1.1. *Fo and Initial Consonants*

It has been widely agreed that there is correlation between Fo on the vowel and the state of voicing of preceding consonants. As summarized in Lehiste (1970:71), "high fundamental frequencies occurred after a voiceless consonant and considerably lower fundamental frequencies occurred after a voiced consonant." This fact has been attested in various languages such as English (House and Fairbanks 1953, Lehiste and Peterson 1961, among others), French (Fischer-Jørgensen 1972), Swedish (Löfqvist 1975), Yoruba (Hombert 1976), etc. However, there are conflicting data regarding the effect of aspirated versus unaspirated voiceless stops on the Fo onset of a following vowel. Some studies show that voiceless aspirated stops give rise to a higher Fo onset, as in Korean (Han 1967, Kim 1968), Thai (Ewan 1976), Danish (Jeel 1975). Others show no difference in Fo onset, as in Danish (Fischer-Jørgensen 1968), French and English (Hombert and Ladefoged 1977). And others still show the opposite pattern, i.e. higher Fo onset after voiceless unaspirated stops, as in Korean (Kagaya 1974), Thai (Gandour 1974, Erickson 1975) and Hindi (Kagaya and Hirose 1975).

Another interesting fact was observed by Lea (1973:61-62) following a study of English word pairs with contrasting stress patterns; he noted that "falling or rising Fo contours at vowel onset do not simply mark either stress or state of voicing, but are functions of a complex combination of stress and phonetic context," or more precisely, "a rising Fo contour at vowel onset may indicate either that the vowel is word-initial or preceded by a voiced consonant, or that the syllable is stressed (with either a voiced <u>or</u> unvoiced prevocalic consonant."

Also interesting are some observations made in Hombert (1978:83) from Yoruba data that "the perturbation caused by a voiced consonant on a following high tone or by a voiceless consonant on a following low tone is greater than the effect of these two series of consonants on a mid tone," and his remark that "there may be a tendency in tone languages (which does not exist in non-tonal languages) to minimize the intrinsic effect of prevocalic consonants actively - probably in order to render the different tones maximally perceptually distinct."

My data, given in Tables 3.1 and 3.2 (next pages), tend to support the above-mentioned observations. Table 3.1 gives the mean Fo onset values and standard deviations in Hertz of the same vowel /a/ after different consonants. Levels of significance obtained from two-tailed t-tests on the means show that there is significant difference between Fo onsets after voiceless and voiced consonants as two separate classes, although the voiceless /t/ sometimes gives rise to Fo onsets closer to those of /d/ than of /t^h/. Table 3.2 gives the mean Fo values at various timepoints for two tones, level and falling, on syllables with voiceless and voiced initial consonants, by four female informants. Figure 3.1 (p 138), where the data from Table 3.2 are plotted, serves to illustrate the notions of Fo loci and pitch targets which I propose as abstract constructs to account for those facts.

Leaving aside, for the time being, the physiological causes that might determine them, let us define Fo loci as an imaginary Fo range for certain classes of sounds and pitch target as an ideal pitch level for certain tones or intonation. I use the term 'Fo loci' to mean that the Fo levels toward which different segment types tend to reach are intrinsically determined by articulatory or phonatory factors, e.g. laryngeal tension, associated with their production, which the speaker do not control intentionally. I use the term 'pitch target' to mean that what the speaker intends

TABLE 3.1 Mean Fo Onset Values in Hertz of Same Vowels in Vietnamese Syllables With Different Initial Consonants [t], [t^h], [s], [d] and [n]

Informant	n	Mean Fo Onset and SD After [t]	[t^h]	[s]	[d]	[n]	C_{vl}	C_{vd}	Significance of Difference Between Means t/t^h	t/d	s/d	t^h/d	t^h/n	C_{vl}/C_{vd}
NF1, NF2, CF1, CF4, SF1, SF5	48[(a)]	244 35	256 39	254 36	232 40	237 41	251 37	234 41**	ns*	ns	0.01	0.01	0.05	0.01
NF1	8[(a)]	222 12	240 14	238 18	213 10	221 6	233 14	217 6	0.05	ns	0.01	0.01	0.05	0.01
NM4	10[(b)]	110 4	119 4	111 5	106 4	106 4	113 5	106 4	0.01	ns	0.05	0.01	0.01	0.01

* not significant at the 0.05 level. ** Mean Fo on first line, SD on second line.

(a) For each informant, eight repetitions, with different tones, of syllables with same vowel [a] and beginning with [t], [t^h], [s], [d] and [n]. The numbers for C_{vl} and C_{vd} are the totals of [t,t^h,s] and [d,n] respectively.

(b) Ten repetitions of syllables as in (a) with the same rising tone /ʹ/.

TABLE 3.2 Mean Fo Values in Hertz of Level and Falling Tones with Voiceless and Voiced Initial Consonants by Four NV and SV Informants (NF1, NF2, SF1 and SF5)

Tone	Initial Consonant	n	Mean Fo and SD at Timepoint						
			CO	P1	P2	P3	P4	P5	P6
Level	Voiceless	12		280	278	274	271	271	265
	[t,t^h,s]			19	18	21	23	23	25*
	Voiced	8	252**	265	267	270	266	263	261
	[d,n]		22	26	24	22	23	26	27
Falling	Voiceless	12		242	231	224	215	212	217
	[t,t^h,s]			27	30	29	29	27	27
	Voiced	8	218	224	223	221	216	215	218
	[d,n]		22	20	21	20	23	26	22

* Mean Fo on first line, SD on second line.

** CO values were measured at first stable point of consonant onset for [d] and [n].

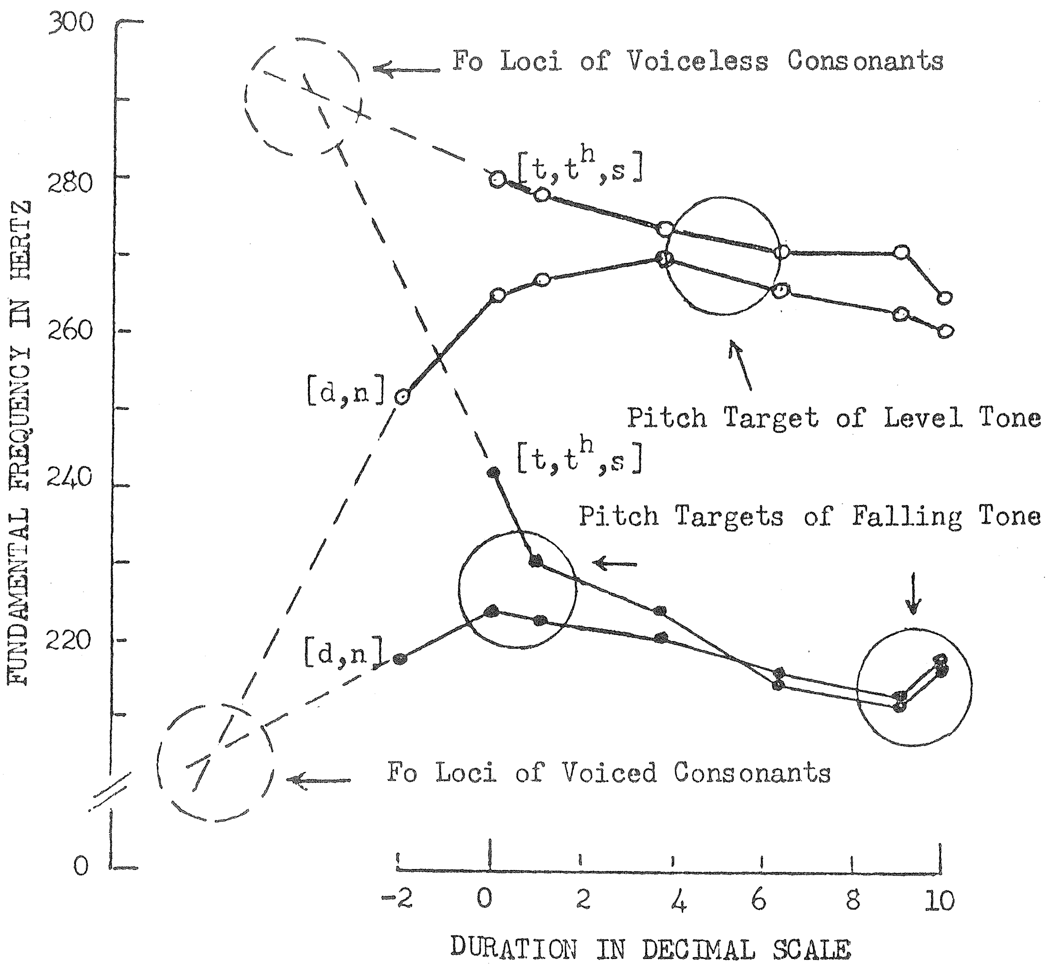


FIGURE 3.1 Mean Fo of Syllables with Level and Falling Tones with Voiceless and Voiced Initial Consonants Plotted Against Normalized Duration (Data from Table 3.2)

Fo values at timepoint -2 represent the mean Fo onsets of syllables with voiced initial consonants [d] and [n]. Dashed lines extending from the onsets meet at their imaginary Fo loci represented by dashed circles. Solid-line circles represent the tonal pitch targets.

and extrinsically controls is relative pitch levels within his own voice range and which he can monitor during their production. It is the dynamic interaction between these intrinsic and extrinsic factors that produces perturbations in Fo realizations. The degree of perturbation will depend on the amount of adjustment needed to accommodate them. Thus there is more perturbation when a consonant with high Fo loci begins a tone with initial low target than when it begins a tone with initial high target, etc. I have preferred the term 'loci' and 'target' to others because they represent something to be reached but not always realized in actual speech because of numerous influences.

As illustrated in Figure 3.1, voiceless consonants have high Fo loci and voiced consonants have low ones, which could be located in the upper and lower reaches of a speaker's average voice range, respectively. In the figure, they could be extrapolated by extending imaginary lines from the onsets of the Fo curves of different tones preceded by the same type of consonants and the areas where these would meet constitute the Fo loci. My illustrative examples are based on only two tones and five different consonants. More data on other tones and other consonants would be needed to support the hypothesis.

The tonal pitch targets for Vietnamese can be determined from the typical Fo contours of NV, CV and SV tones described in Ch. Two, and represented by the system of tone letters proposed in Chao 1930. Formal processes for deriving them from the acoustic material will be proposed in Ch. Five, but they could be roughly obtained from auditory impressions or visual inspection of the

Fo curves. Thus, the level tone has one target at midpoint, the rising and falling tones have two, at onset and endpoint, and the concave tones have three, at onset, midpoint and endpoint. In this respect, I propose the consistent use of one, two or three figures for different tones according to the number of pitch targets they represent, e.g. [3] (and not [33]) for the level tone, [35] for the rising tone, [325] for the broken tone, etc.

These concepts could account for most of the facts observed at Fo onsets of different syllables, as referred to earlier in this sub-section. They do not explain how and why Fo is affected in such manner, a question we shall examine later.

3.1.2 Fo, Duration and Final Consonants

Earlier studies reveal no significant correlations between Fo and following consonants. As summarized in Lehiste (1970:74), "final consonants showed no regular influence on the fundamental frequency of the syllable nuclei." However, some studies (Mohr 1968, Slis 1966) indicated that "post-vocalic consonants have an effect on Fo similar to that of pre-vocalic consonants, but with a much smaller magnitude," as quoted in Hombert 1978. This latter study also reported that post-vocalic /h/ and /ʔ/ produced significant perturbations on the Fo of the preceding vowel in Arabic, /h/ lowering and /ʔ/ raising the Fo perceptibly. Hombert, Ohala and Ewan 1979 suggested that "if a tonal contour appears on a vowel which is followed by a voiceless consonant, it may be 'cut short' and have a different terminal Fo than if it appears before a voiced consonant." And it is a well known fact that vowels are shorter before voiceless than before voiced consonants (Delattre 1962, Chen 1970, Peterson and Lehiste 1960, etc.)

The data on Vietnamese tones reveal interesting facts in this respect. First, concerning duration, Han 1969 noted that for some of her informants the creaky-ending NV drop tone is about half - as long as the other tones (stopped tones not included.) Similar results were obtained by Earle 1975 who reported that on the average the creaky-ending NV drop tone is 47% shorter than the mean duration of other tones (data from eleven NV speakers.) He also reported that the non-stopped and stopped NV drop tones are of the same duration, but the stopped rising tone is only about 25% as long as the corresponding non-stopped rising tone. My data show more consistent duration ratios across the three dialects for the stopped tones : about 60% of the average longer tones, while the NV creaky-ending drop tone is 80% and the CV creaky-ending curve tone is 90% of the longer tones' duration. In spite of the discrepancies between the above-mentioned studies, which may have arisen from different measurement procedures or from the fact that the data involved different types of syllables, one may conclude that Vietnamese syllables ending in a voiceless stop or creaky voice are regularly shorter than the average. Thus, duration is predictable from syllable types and is not a feature of tone in Vietnamese.

Second, there is a tendency for the stopped variants of the rising tone to rise higher than the corresponding non-stopped tone, while there is no such tendency for the stopped variant of the drop tone. The data in Earle 1975 show that the drop tone and its stopped variant exhibit no statistically significant difference in Fo or duration, but the stopped rising tone is not only much shorter but its Fo matches the second half of the non-stopped rising tone, i.e. its onset is much higher. Data from three of my informants, one from each dialect, are presented in Table 3.3 and Figure 3.2 (next pages) for a better illustration of the patterns which are already apparent in the general description of the tones in Ch. Two. Two-tailed t-tests on means show that there are significant differences between the mean Fo onsets and endpoints of the non-stopped and stopped rising tones by all three informants, while the differences are not significant between the mean Fo endpoints of the NV drop tone and its stopped variant, and between the mean Fo onsets of the CV and SV drop tones and their variants.

The above facts lend further support to the idea of interaction between Fo loci and pitch targets. Although the Fo loci of final stops do exert some influence on the Fo of preceding vowels, their effect is not the same in every case because of interaction with pitch targets. The process is schematized in Figure 3.3 (page 145) where the Fo loci of final stops and tonal pitch targets are shown to combine with each other to give greater effect or to neutralize each other so that no appreciable effect is produced.

Thus, for example, the non-stopped rising tone has two pitch targets and there is pitch movement between them, represented as 3↗5. The corresponding stopped tone can be supposed to have the same underlying contour, but it is followed by a voiceless stop which has high Fo loci whose effect on the Fo of the preceding vowel can be represented as ↑C. The interaction between targets and loci can be represented as 3↗5 ↑C and the predicted result is 35↗ because

TABLE 3.3 Mean Fo Values in Hertz of Rising and Drop Tones and Their Stopped Variants by Three Informants

Informant	Tone	n	Mean Fo and SD at Timepoint P1	P2	P3	P4	P5	P6
NF2	Rising /ʹ/	8*	275	267	269	286	307	321
			19	14	8	11	27	35**
	St. Ris. /ʹs/	8	318	327	340	354		
			16	15	11	9		
	Drop /./	8	246	243	229	211	202	208
			14	11	14	15	25	24
	St.Drop /.s/	8	234	216	206	199		
			18	20	13	10		
CM2	Rising /ʹ/	8	124	121	124	134	142	146
			7	3	4	6	5	5
	St. Ris. /ʹs/	8	132	138	147	157		
			12	11	7	6		
	Drop /./	8	122	119	114	107	104	104
			7	4	4	3	4	3
	St. Drop /.s/	8	124	121	118	112		
			7	3	2	7		
SF1	Rising /ʹ/	8	304	296	302	326	353	364
			19	13	9	11	15	19
	St. Ris. /ʹs/	8	328	327	356	391		
			21	8	13	14		
	Drop /./	8	281	264	252	259	269	268
			9	4	11	5	7	8
	St. Drop /.s/	8	277	264	256	254		
			15	6	4	7		

* Eight syllables with same vowel [a] preceded by different consonants, and followed by [k] in the case of stopped tones.

** Fo values on first line, SD on second line.

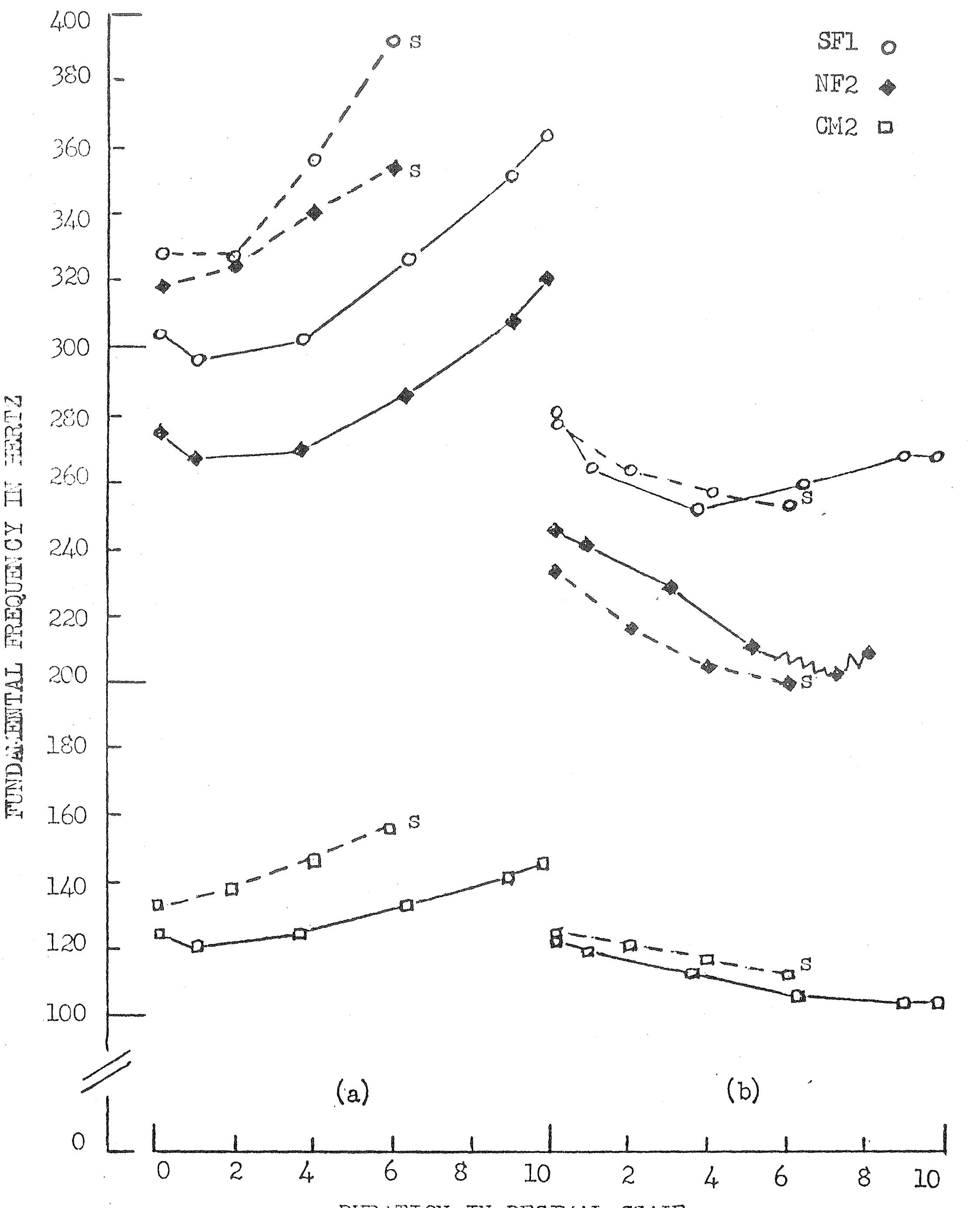


FIGURE 3.2 Mean Fo of (a) Rising Tones and (b) Drop Tones with Their Stopped Variants (Dashed Lines) by Three Informants Plotted Against Normalized Duration (Data from Table 3.3)

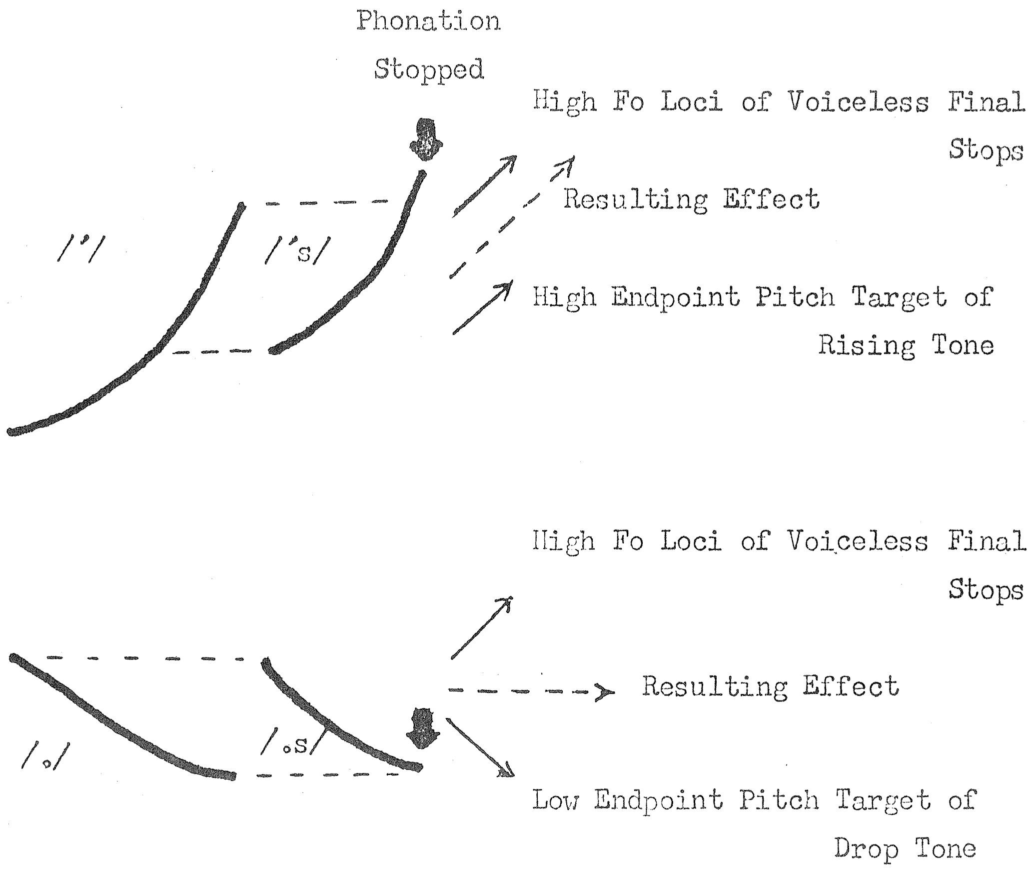


FIGURE 3.3 Schematic Representation of the Combined Effect of Final Stop Fo Loci and Tonal Pitch Targets on the Fo of Vietnamese Stopped Tones.

Light arrows represent movements toward intrinsic Fo loci of consonants and intended pitch targets of tone; dashed arrows represent resulting effect; heavy arrows represent the stopping of phonation which shortens the tones.

the pitch movement toward the final target is reinforced by the 'pull' of high Fo loci of a voiceless stop in the same direction. This type of result is illustrated by the significantly higher Fo values of the stopped rising tone in all three dialects shown in Table 3.3 and Figure 3.2.

In the case of the stopped drop tone, the interaction can be represented as 2↘1 ↑C. The predicted result can be represented as 21↘↑--→ ? , i.e. no appreciable perturbation or variable perturbation on the basic Fo contour. Because of opposing influences, the final result may depend on which of them is stronger. This type of result is illustrated in the Fo values of the stopped drop tones in Table 3.3 and Figure 3.2, where no significant differences or differences in opposite directions with the non-stopped drop tones are found.

3.1.3 Vowel Quality and Fo

It has been widely attested that there is correlation between vowel 'height' and Fo, i.e. high vowels (high tongue body, low F1) such as /i/, /u/ have intrinsically higher Fo than low vowels (low tongue body, high F1) such as /a/, /ɔ/, etc. The fact is found in various languages such as English (Peterson and Barney 1952, House and Fairbanks 1953, etc), Serbo-Croatian (Ivic and Lehiste 1963), Korean (Kim 1968), Vietnamese (Han 1969), Danish (Petersen 1976), among others. The best known theories proposed to explain this phenomenon are the source-tract coupling theory (Flanagan 1965, Flanagan and Landgraft 1968, Lieberman 1970, etc.) and the tongue-pull theory (Ladefoged 1964, Lehiste 1970, Ohala

1972, etc.), neither of which has received unqualified empirical support for their predictions (Hombert 1978).

The problem that interests us here is the relationship between intrinsic vowel Fo and tonal pitch targets. For comparison, data on NV from Han 1969 and on Taiwanese from Zee 1978 are given in Table 3.4 and graphically represented in Figures 3.4 and 3.5 (following pages). The Vietnamese data come from two informants, a male and a female, and include 30 occurrences of each of the vowels /i/, /e/, /a/, /o/ and /u/, with each of the three tones (level, falling, and rising) which were measured at one, two and three data points respectively. The Taiwanese data come from three male informants and include the means for five occurrences of the five vowels with two tones, computed from values at 10 milliseconds intervals.

For the Taiwanese data, Zee 1978 concluded that "the value of the fundamental frequency of the high [level] tone correlates with vowel height for all three speakers...No such correlation is found in vowels associated with the low [falling] tone for all three speakers." The Vietnamese data from the female Informant F agree with the above conclusion, but those from the male Informant M exhibit an opposite pattern: correlation in the case of the falling tone, no correlation in the case of the rising tone. A careful examination of Figure 3.5 where Vietnamese tones are represented not in average Fo for the whole vowels but Fo at different data points, would suggest that the relationship between vowel height and Fo is more complex with each tone. There seems to be more correlation between vowel height and Fo at the onset of the rising tone than at its endpoint for both informants, and more correlation

TABLE 3.4 Mean Fo of Vowels with Different Tones
(NV Data from Han 1969; Taiwanese Data from Zee 1978)

Language & Informant	Tone	Data Points	n	Mean Fo in Hertz for				
				i	e	a	o	u
NV F(emale)	Rising	Onset	30	281	258	254	258	281
		Mid	30	281	258	254	258	281
		End	30	357	360	348	359	399
		$\bar{F}$*	90	306	292	285	291	320
	Level	Mid	30	309	293	288	288	308
	Falling	Onset	30	227	224	224	216	221
		End	30	189	190	193	185	184
		$\bar{F}$	60	208	207	209	201	203
M(ale)	Rising	Onset	30	137	130	130	131	135
		Mid	30	139	135	134	134	135
		End	30	156	156	156	155	156
		$\bar{F}$	90	144	140	140	140	142
	Level	Mid	30	141	139	136	137	143
	Falling	Onset	30	146	136	128	139	142
		End	30	117	109	106	115	116
		$\bar{F}$	60	132	122	117	127	129
Taiwanese								
A (male)	High	$\bar{F}$	5**	134	124	121	133	137
	Low	$\bar{F}$	5	81	77	79	84	84
B (male)	High	$\bar{F}$	5	155	146	142	149	157
	Low	$\bar{F}$	5	106	103	106	104	105
C (male)	High	$\bar{F}$	5	181	167	166	170	176
	Low	$\bar{F}$	5	110	117	117	115	116

* Mean Fo for all datapoints.

** 5 tokens for each tone; for each token, the mean is taken of Fo values at 10 msec intervals.

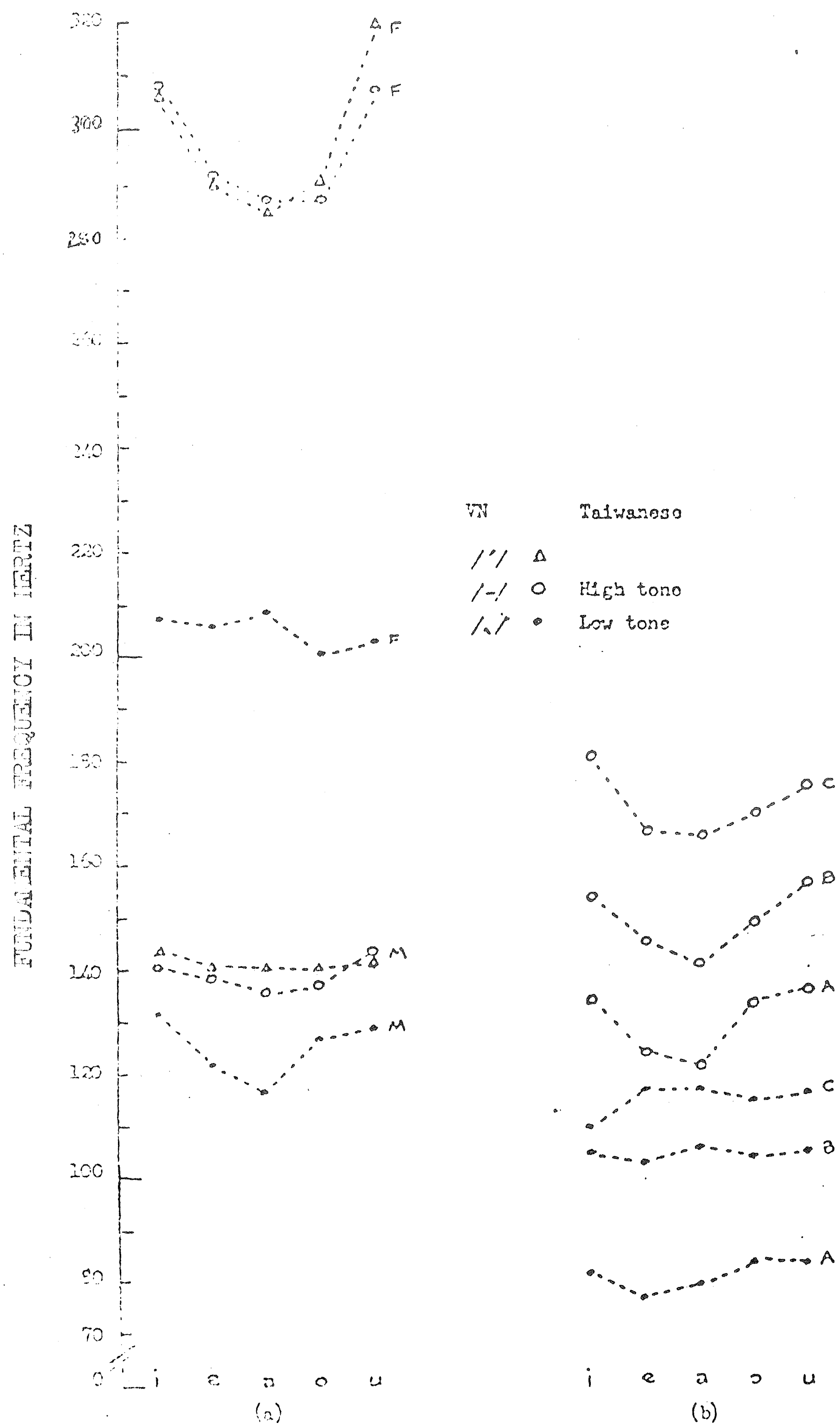


FIGURE 3.4 Mean Fo of (a) Vietnamese and (b) Taiwanese Vowels Associated with High and Low Tones (Data from Han 1969 and Zee 1978. The discontinued lines join the mean Fo of five typical vowels)

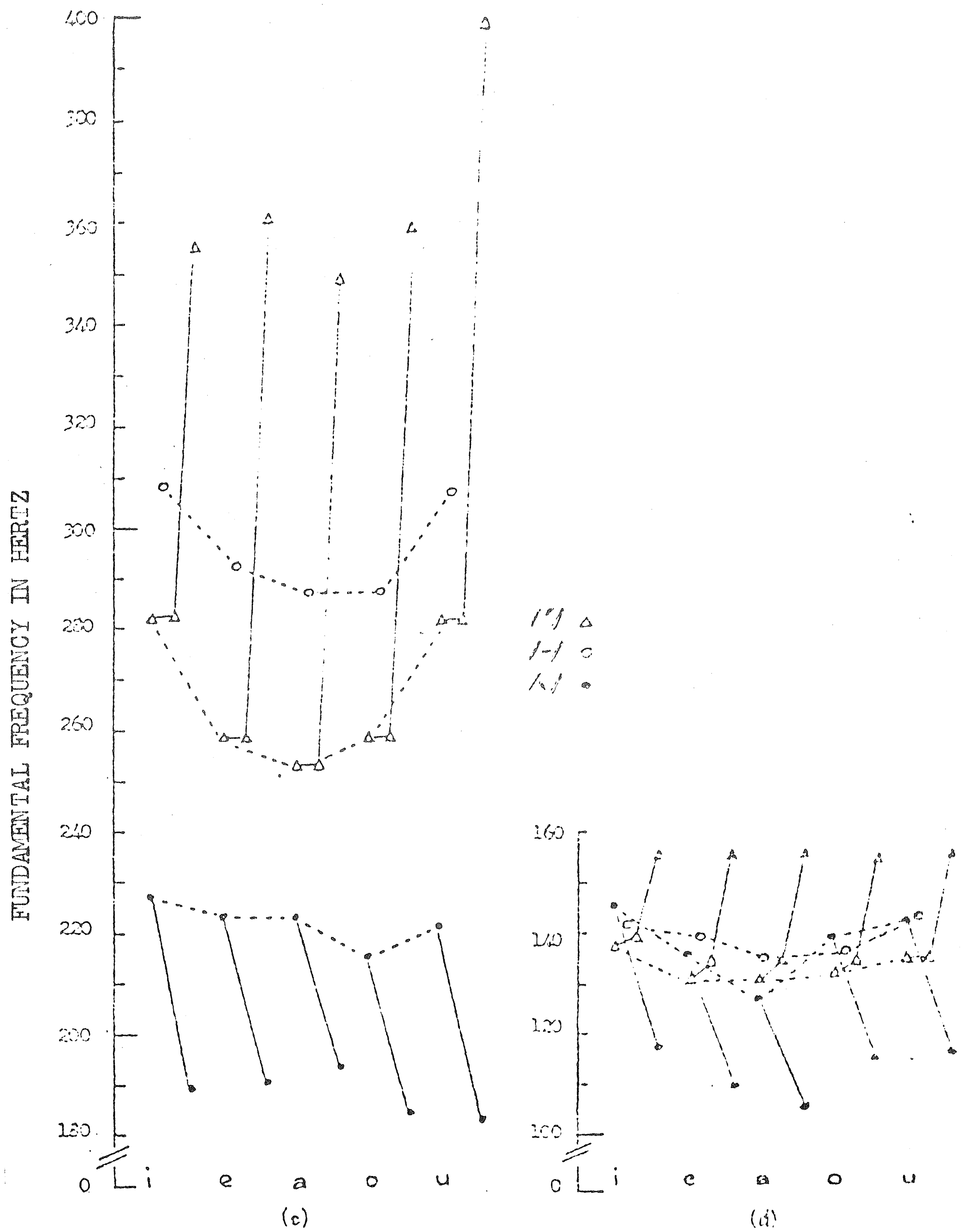


FIGURE 3.5 Mean Fo of NV Vowels by (c) Male Informant M and (d) Female Informant F, Associated with Different Tones (Data from Han 1969. The discontinued lines join the Fo onsets of five typical vowels. The continuous lines join Fo at different data points of the same tone on the same vowel)

when the Fo onsets of the falling tone by Informant M are in the middle ranges of his voice than those of Informant F when they are much lower within her voice range. The data suggest that correlation between vowel height and Fo is usually disturbed when the pitch targets require Fo movements toward the more extreme voice ranges, and this may vary with each tone and individual speakers.

3.1.4 Fo and Tonal Environment

Fo variations due to tonal environment are usually described as tone sandhi or other forms of tone assimilation in various languages, e.g. various Chinese dialects (Kratochvil 1968, Chang 1958, Wang 1967, Yip 1980, among others) and many African languages (Hyman 1973b, Leben 1973, 1978, Schuh 1978, among others).

Concerning Vietnamese, Han and Kim 1972 (p 62) noted that "the phonetic variation of the six tones never [my emphasis] leads to a phonetic overlap between any two tones in an identical environment," and Tran Huong Mai 1969 (p 283) made similar but stronger claims, asserting that "a tone in Vietnamese is never replaced by another phonemic tone in any environment...The relative pitch contrast between the five [SV] tones is always maintained and the opposition between them is never neutralized [my emphases]."

While agreeing that tone sandhi hardly exists in Vietnamese in the usual sense, possibly due to the optimal differentiation of tones in pitch contour, pitch level and laryngealization, I would replace the word never by rarely in the above quotations to make them more realistic. The data presented below will show that Fo perturbations due to tonal environment are sometimes considerable in terms of absolute Fo values and they occasionally give rise to tone shapes quite remote from the standard forms.

The first set of data are taken from Han and Kim 1972. Table 3.5 (next page) presents the differences between the mean Fo of the highest and lowest variants of some typical tones in different tonal environments. They occur in two-syllable utterances by a female NV informant, and the Fo was measured at three data points (onset, midpoint and endpoint). The upper part of the table shows the highest and lowest variants in second syllable position, and the lower part shows them in first syllable position. One may note that (a) in both positions, the highest variants occurred after or before a high tone (/´/, /~/ or /-/), and the lowest variants occurred after or before a low tone (/\/ or /ʾ/); (b) the differences between the highest and lowest variants are usually greater at onset than at endpoint; and (c) they are greater when the tones are in second syllable position than in first syllable position. Han and Kim 1972 inferred from the above facts that "the progressive effect, which is defined as the effect of a tone on another in the immediately following syllable, is greater than the regressive effect which is the effect in the opposite direction," and that "the endpoint of the tones is more stable and resistant to the environmental influence."

The second set of data are based on test-words bearing different tones from the following list:

/zāj/	'tenacious'	/zàj/	'long'
/zój/	'carelessly done'	/zót/	'ignorant'
/zạj/	'foolish'	/zẹt/	'flattened'

TABLE 3.5 Differences Between Mean Fo of Highest and Lowest Variants of Same Tones in Different Tonal Environments (Data from one NV female informant, in Han & Kim 1972)

Syllable Position	Tone	Environment	n	Mean Fo of Highest Variant at			Environment	n	Mean Fo of Lowest Variant at			Difference in Hertz Between	
				O_1	M_1	E_1 *			O_2	M_2	E_2 *	O_1-O_2	E_1-E_2
Second**	Rising	After /'/	72	294	290	367	After /'/	30	190	218	309	104	58
	Broken	- /~/	42	282	? **	406	- /'/	18	193	?	342	89	64
	Falling	- /'/	57	243	215	194	- /̥/	40	203	200	174	40	20
	Curve	- /~/	6	243	157	157	- /'/	36	184	151	178	59	-21**
First**	Rising	Before /'/	72	245	259	311	Before /̥/	57	210	226	295	35	16
	Level	- /~/	72	284	301	295	- /̥/	72	236	249	250	48	45
	Falling	- /-/	63	223	224	221	- /̥/	48	206	208	202	17	19
	Curve	- /-/	54	194	167	161	- /'/	36	184	157	166	10	-5

* Onset, Mid and End values in Hertz respectively.

** Second and first in two-syllable utterances of various tone combinations.

** ? means laryngealized, no Fo values obtained; a negative value means E_2 has higher Fo than E_1 .

/zɤ̉/ 'poor (quality)' /zẽ/ 'easy'

They were placed in two frame sentences

(a) /kũŋ xá --- dɤ́j/

also fairly --- there 'it is fairly ---'

(b) /--- t^hɤ̣̆t là ---/

--- really be --- 'it is really ---'

The forms /dɤ́j/ and /t^hɤ̣̆t/ were alternatively realized as /dɔ́/ and /t^hiə̣t/ or /t^hiə̣k/ according to dialect usage. The frames provided environments for the tones in test-words to occur (a) after and before a rising tone, (b) in utterance-initial position and before a drop tone, and (c) before a falling tone and in utterance-final position.

Tables 3.6 and 3.7 and Figure 3.6 (next pages) present the data from various NV, CV and SV informants. Table 3.6 shows that the differences between mean Fo onsets of the tones preceded by a rising tone and those preceded by a falling tone are all significant, but those in utterance-initial position are not always significantly different from the two other cases. Table 3.7 shows that there are generally no significant differences at endpoint when the tones are followed by either the rising tone, the falling tone or in utterance-final position. Figure 3.6 illustrates the pattern with data from one informant, SM5.

The two sets of data display the same tendency of Vietnamese tones to be more affected by tonal environment at onset than at endpoint. A possible explanation for that can be found in the structure of the Vietnamese tones itself. The endpoint targets are more likely to be realized than the onsets because (a) they are usually at

TABLE 3.6 Differences Between Mean Onset Fo of Syllables Preceded by A Rising Tone, A Falling Tone and in Utterance-Initial Position (Data from NV, CV and SV Informants)

Informant	n	Mean Fo(O_1) /'/ $\bar{x}$	SD	Mean Fo(O_2) /ˎ/ $\bar{x}$	SD	Mean Fo(O_3) Ø $\bar{x}$	SD	Difference in Hz and in % (and Level of Significance) O_1-O_2			O_1-O_3			O_2-O_3		
NF3, NF4	16	268*	17	211	15	228	19	57	27%	(0.01)	40	17%	(0.01)	17	-8%**	(0.02)
NM1, NM4	16	133	13	114	8	120	10	19	17%	(0.01)	13	11%	(0.01)	6	-5%	(ns)**
CF3	8	249	12	222	6	235	8	27	12%	(0.01)	14	6%	(0.05)	13	-6%	(0.01)
CM2, CM3, CM6	24	124	9	114	6	119	6	10	9%	(0.01)	5	4%	(0.05)	5	4%	(0.01)
SF3	8	236	31	199	14	220	21	37	18%	(0.02)	16	7%	(ns)	21	-10%	(0.05)
SM2, SM5, SM6	24	128	12	116	12	121	12	12	10%	(0.01)	7	6%	(ns)	5	4%	(ns)

* Means taken of Fo onset values of different tones in same environments.

** Percentages rounded to the nearest unit; a negative value means O_3 has higher Fo than O_2.

** (ns) means the difference is not significant at the 0.05 level.

TABLE 3.7 Differences Between Mean Endpoint Fo of Syllables Followed by A Rising Tone, A Falling Tone and in Utterance-Final Position (Data from MV, CV and SV Informants)

Informant	End-pts	n	Mean Fo(E_1) /'/ $\bar{x}$	SD	Mean Fo(E_2) /\/ $\bar{x}$	SD	Mean Fo(E_3) Ø $\bar{x}$	SD	Difference in Hz and in % (and Level of Significance) E_1-E_2			E_1-E_3			E_2-E_3		
MF3, MF4	H*	6	309	26	304	18	335	25	5	2%	(ns)	26	-8%	(ns)	31	-10%	(0.05)**
	L	6	215	20	198	14	179	14	17	8%	(ns)	36	20%	(0.01)	19	10%	(ns)
MM1, MM4	H	6	172	16	170	17	190	18	2	1%	(ns)	18	-10%	(ns)	20	-12%	(ns)
	L	6	108	12	105	11	94	4	3	3%	(ns)	14	14%	(0.05)	11	12%	(ns)
CF3	H	3	247	4	250	8	246	4	3	-1%	(ns)	1	0.4%	(ns)	4	2%	(ns)
	L	3	225	7	213	4	213	6	12	6%	(0.05)	12	6%	(ns)	0	0	(ns)
CM2, CM3	H	9	133	16	135	14	127	13	2	-1%	(ns)	6	5%	(ns)	8	6%	(ns)
CM6	L	9	110	5	105	6	104	5	5	5%	(ns)	6	6%	(ns)	1	1%	(ns)
SF3	H	3	316	32	358	31	341	31	42	-13%	(ns)	25	-8%	(ns)	17	5%	(ns)
	L	3	195	7	203	4	195	7	8	-4%	(ns)	0	0	(ns)	8	4%	(ns)
SM2, SM5	H	9	161	9	167	14	171	20	6	-4%	(ns)	10	-6%	(ns)	4	-2%	(ns)
SM6	L	9	107	11	110	11	110	14	3	-3%	(ns)	3	-3%	(ns)	0	0	(ns)

* Means taken separately of high endpoint Fo (H) and low endpoint Fo (L) of similar rising or falling tones.

** Percentages rounded to the nearest unit; a negative value means the second member of the comparison has higher Fo than the first; (ns) means the difference is not significant at the 0.05 level.

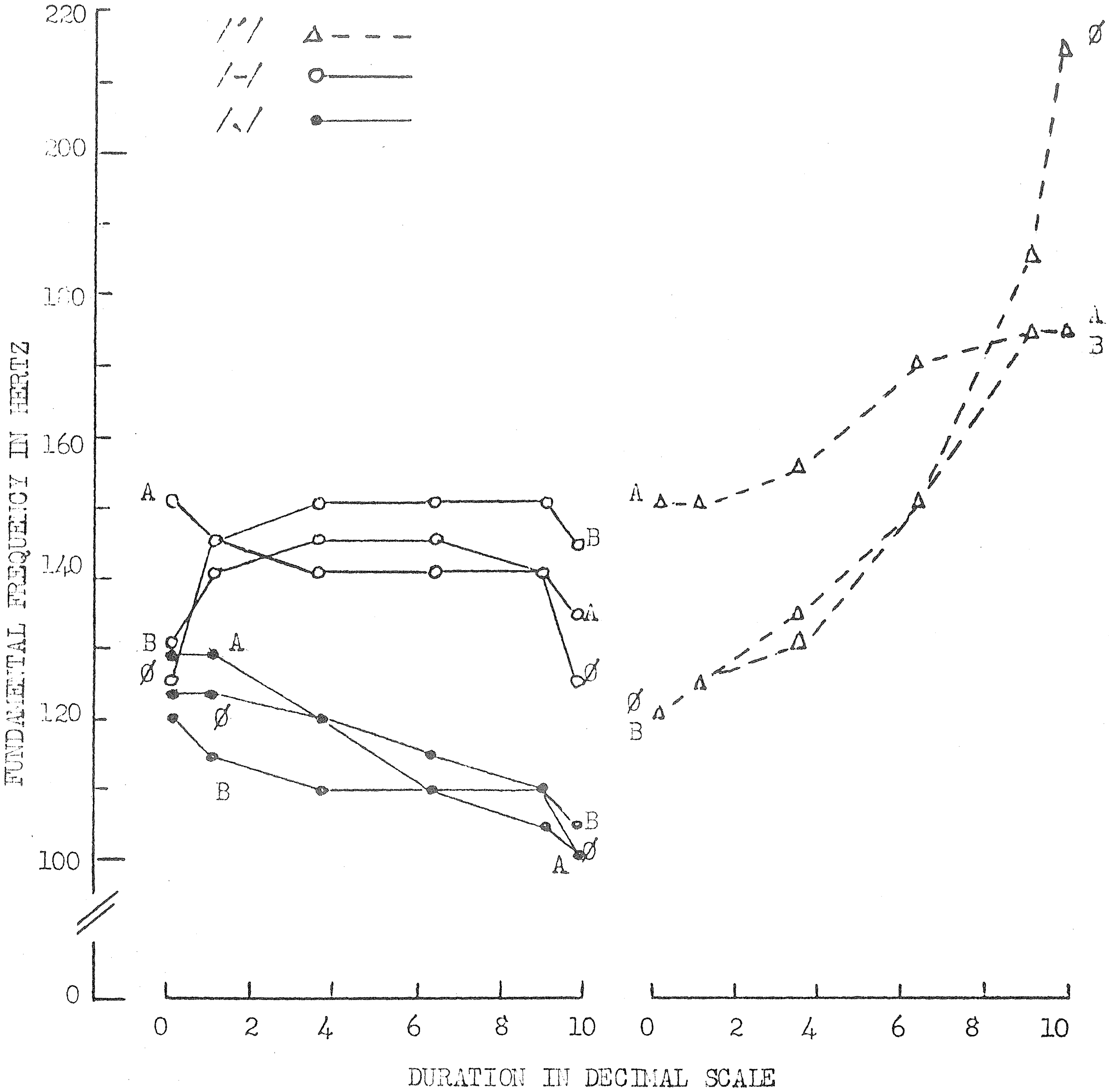


FIGURE 3.6 Fo of Three Different Occurrences of SV Level, Falling and Rising Tones by Informant SM5 Plotted Against Normalized Duration

A – A: Preceded and Followed by Rising Tone

B – Ø: Preceded by Falling Tone and in Utterance Final Position

Ø – B: In Utterance Initial Position and Followed by Drop Tone extreme Fo ranges and may need greater muscular or aerodynamic efforts, therefore less subject to coarticulation pressure than the onsets which usually occur at or near the average Fo ranges; (b) tonal contour, which has been shown to have primary importance in the Vietnamese system, allows that the onset may be started at greatly different pitch levels, but requires that movements toward the end-point targets be tone-specific to preserve tonal contrast. The same reasons may account for the fact that there is little phonetic or phonemic overlap between Vietnamese tones as noted earlier. There are cases when such overlap occurs, but they are rare.

The only instances of regular tone sandhi at the phonemic level in Vietnamese seems to occur in some reduplicative patterns described in various studies (Nguyen Tai Can 1975, Ho Le 1976, Dai Xuan Ninh 1978, among others). They involve some unstressed reduplicated syllables which change to an 'even' tone (level or falling) belonging to the same pitch register than the basic syllables, which may be of any tones. For example

/hẵj/ 'interesting' /hă̄j hă̄̃j/ 'rather interesting'

but /ʈắŋ/ 'white' /ʈă̄ŋ ʈắŋ/ 'whitish'

/nặŋ/ 'heavy/ /nằŋ nặŋ/ 'rather heavy'

Otherwise, Fo variation usually operates at the subphonemic level. There would be no need for tone sandhi rules in Vietnamese except those involving the reduplicative patterns. If one tries to describe subphonemic tone variations in speech, the rules would be extremely complex because there are many factors that may interact with the basic pitch targets to produce variations in Fo contours and Fo levels. They may be inherent in the syllable structure, e.g. consonant types and vowel quality. They may also come from other suprasegmental, extrasyllabic or even extralinguistic elements, e.g. stress, intonation, adjacent tones, syllable position in the utterances, constituent structure (Lea 1973),emotion, rythm and tempo (Crystal 1969).

3.2 DYNAMIC INTERACTION IN THE PHONATION PROCESS

The facts observed and described in the previous section point to regular correlations between Fo and initial consonants, Fo and final consonants, Fo and vowel quality and between the Fo at onset and endpoint of adjacent tones.

They may be supposed to reflect the dynamic interaction between various intrinsic factors and the basic pitch targets of a tone, which represent its idealized pitch contour as intended by speakers. To understand the physical and physiological nature of this interaction which presumably occurs at the laryngeal level, I shall look at the mechanisms that produce voice and pitch in speech and speculate on some causal relationships.

It has been widely accepted, as stated in Ladefoged 1971 (p 7) that "the rate of vibration during a voiced sound depends on two factors: the tension of the vocal cords and the pressure drop across them." However, the occurrence of creaky voice or heavy laryngealization as a characteristic of some Vietnamese tones and the correlations mentioned earlier suggest that there are more complex relationships between the physiological mechanisms that produce voice, pitch and various phonation types.

Explaining the aerodynamic-myoelastic theory of phonation, van den Berg 1968 stated that

> the function of the larynx is based on the interplay of three factors: (1) the aerodynamic properties of the air which actuates the larynx, (2) the adjustment of the larynx, brought about by the proper nervous activation of the various muscles, and the myoelastic properties of the laryngeal components, (3) the aerodynamic coupling between (a) the subglottal system and the larynx, (b) the left and right vocal fold, and (c) the larynx and the supraglottal system. (van den Berg 1968: 291-292)

Ohala 1973, 1978, describing the production of tone, argued that pitch variation is not only caused by the longitudinal vocal cord tension and aerodynamic factors, but also by the vertical tension associated with larynx height. Of the more detailed studies on various aspects of voice quality and laryngeal activity, Catford 1964, 1968 and 1977, Ladefoged 1971, 1975, Hollien 1974 and Laver 1980 gave descriptions of various phonation types commonly found in speech, in terms of their acoustic, aerodynamic, physiological and auditory characteristics. Halle and Stevens 1971 proposed a model that might explain in a unitary way the various facts about pitch, voice quality and segment types. The idea that makes the model attractive is that vocal cord vibration is not controlled by one single factor but is rather the result of a complex interplay of transglottal pressure, glottal opening and vocal cord stiffness, and various laryngeal types can be represented by combinations of these features. However, the model has been criticized in Anderson 1978 and Catford 1977 because the authors make the interacting factors too rigidly linked to each other in the set of features, so that it cannot account, for example, for different combinations of pitch and voice quality.

Concerning Vietnamese tones, the data I have presented permit the following speculations.

(a) Tonal pitch in Vietnamese is presumably regulated by both vocal cord tension and subglottal pressure as described in accepted theories. However, as pointed out in Rose 1981, languages may differ in the mode of control of these parameters. If we consider the characteristics of Vietnamese tones as described in Ch. Two, it appears that Vietnamese gives more prominence to the vocal cord tension factor. The evidence can be found in the data on intensity (which represents sound pressure and is proportional to subglottal pressure). They show that intensity correlates partially with Fo for about two-thirds of duration then usually starts to decay for all tones. Thus the high endpoint Fo of rising tones would be caused by vocal cord tension alone.

(b) The intrinsic high and low Fo loci of voiceless and voiced consonants respectively could have their causes in both the longitudinal tension of the vocal cords and the vertical tension associated with larynx height. Thus initial voiced consonants would be associated with low vocal cord tension and low larynx, both presumably needed to facilitate voicing in the case of low subglottal pressure (caused by increased supraglottal pressure after oral closure for the consonants). Initial voiceless consonants would be associalted with both high vocal cord tension and high larynx, needed to inhibit voicing, hence their raising effect on the Fo of following vowels. Final voiceless consonants have the same effect on preceding vowels when they end a stopped rising tone, because in such a case both high vocal cord tension and high larynx are sustained. However, when they end a stopped drop tone, the low vocal cord tension needed for the low endpoint target would prevail and neutralize the effect of high larynx and high tension needed for the voiceless consonant.

There is also the possibility that high vocal cord tension and high larynx, associated with voicess final stops, favouring voicelessness and inhibiting voicing, are responsible for the shorter duration of stopped tones in Vietnamese. This seems to be related to the fact that vowels are shorter before voiceless consonants than before voiced consonants, and high vowels are usually shorter than low vowels, as observed in various languages (Lehiste 1970).

As for the causes of intrinsic Fo differences associated with vowel quality, the data presented above are not sufficient to be interpreted in favour of either the tongue-pull theory or the source-tract coupling theory.

(c) The occurrence of breathy voice, creaky voice and glottal closure as characteristic or optional features of some Vietnamese tones, as described in Ch. Two, deserves further examination. As noted earlier, these phonation types occur as the results of the complex interplay between aerodynamic factors (subglottal pressure, transglottal airflow) and muscular adjustments (longitudinal tension, adductive tension and medial compression, involving the vocal cords and various laryngeal muscles, according to Laver 1980 and others). Glottal configurations for these types are schematically represented in Figure 3.7 (next page) where dashed lines suggest vibrating cords and thickened lines suggest varying degrees of tension or constriction. The schemata are based on photographs of the glottis taken with a laryngoscope by John Ohala and Ralph Vanderslice (reproduced in Ladefoged 1971, 1975) and various figures in Lieberman 1977, Ohala 1978 and Laver 1980. Ladefoged 1971 (p 8) gave brief descriptions for them as follows

Murmur or 'breathy voice' - arytenoids apart, ligamental vocal cords vibrating

Laryngealization or 'creaky voice' - arytenoids tightly together, but a small length of the ligamental vocal cords vibrating

Glottal stop - vocal cords held together

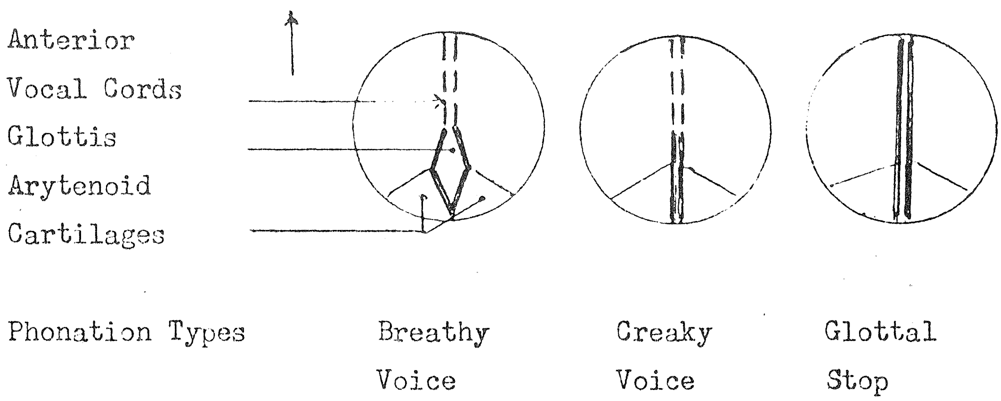


FIGURE 3.7 Glottal Configuration and Phonation Types

However, there can be different realizations of breathy voice and creaky voice. Ladefoged 1975 (p 123) noted that, beside the above description, breathy voice can also be made "with a narrower opening extending over nearly the whole length of the vocal cords, so that when they vibrate they do not actually come completely together, but instead appear to be simply flapping in the airstream." It seems that these two ways of producing breathy voice correspond to what Laver calls 'whispery voice' and 'breathy voice'. Laver noted that "there is a close auditory relationship between breathy voice and whispery voice," but "the physiological relationship between the two is a good deal more distant." (Laver 1980:133). I have used 'breathy voice' in Ladefoged's sense and in keeping with other writings on Vietnamese tones.

Concerning creaky voice, it is also possible that "the ligamental and arytenoid parts of the vocal cords vibrate separately." (Ladefoged 1964, quoted in Laver 1980:139). These possible differences in laryngeal gestures in producing similar auditory effects are consonant with my observations concerning laryngealization in Ch. Two (2.3.6). I suggest that there may be various reasons for such differences : either speakers may use different laryngeal controls when trying to produce the same or similar auditory effects, or the intended laryngeal gestures may be 'overarticulated' or 'underarticulated' by the speakers under certain conditions, or the intended laryngeal gestures may be modified by coarticulation with adjacent segments or other intrinsic factors in various environments.

The foregoing discussion and observations on the phonation process provide possible explanations for tone variation and the interaction of Fo with various intrinsic factors described earlier. The facts suggest that the physical properties of tones, materialized in the acoustic waves, can be understood by referring to the underlying laryngeal activity in which the extrinsically controlled pitch targets intended by speakers interact dynamically with the intrinsic Fo loci of laryngeal types. This might have implications for tonal development in Vietnamese and other South-East Asian languages. They concern the possibility that high tones developed from voiceless initial *consonants, rising tones from final glottal *stops, falling tones from final *h in the proto-languages (as suggested in Haudricourt 1954, 1961, 1972; Matisoff 1973; Hombert, Ohala and Ewan 1979, among others). This may also be related to the development or loss of creaky voice in tonal evolution, as it does not occur on the same tones in various Vietnamese dialects, a point that I have made elsewhere (1980, 1981).

CHAPTER FOUR

CROSS-DIALECT TONE PERCEPTION

4.0 INTRODUCTION

Various studies on the perception of tone have produced similar results in a number of areas. They have shown, among other things, that the tones of real speech syllables in isolation can be readily identified by native speakers of tone languages such as Thai (Abramson 1972, 1975, 1976), Mandarin Chinese (Chuang et al. 1972, Howie 1976), or Cantonese (Fok 1974). Fundamental frequency has been found to be the most important cue for the perception of tone in these studies, and also in Yoruba, an African language (Hombert 1976). Experiments with whispered isolated tones showed that they were poorly recognized by native speakers of Thai (Abramson 1972) and Vietnamese (Miller 1961). On the other hand, additional information on relative intensity may enhance the perception of tone, as shown in synthetic speech experiments with Thai speakers (Abramson 1975). Kratochvil 1971 found that Fo, intensity and duration all contributed to the identification of standard Chinese tones. More recently, Gandour and Harshman 1978 and Gandour 1978 applied the technique of multi-dimensional scaling analysis to determine the relative perceptual importance of various tonal dimensions such as average pitch, direction of pitch movement, length, extreme endpoint and slope. They found that native speakers of English, Thai and Yoruba showed differences in "relative perceptual saliency" of different tonal dimensions and concluded that the way tone is perceived may vary with the listener's language.

We may assume that the same principles apply to the perception of tone in Vietnamese. However, given the characteristics of Vietnamese tones in the three dialects as described in Ch. Two, and the fact that the dialects are mutually intelligible, we may further predict that for Vietnamese, (a) the tones in one dialect can be readily identified by speakers of other dialects with some differences due to their backgrounds; (b) both relative pitch and pitch contour may be important cues for tone perception in all three dialects, and creaky voice an important cue for identifying some NV and CV tones; and (c) in a cross dialect situation, pitch contour may become the single most important cue. This can be inferred from the fact that, of the eight phonetic tones of standard Vietnamese (six underlying tones and two stopped variants), six have the same basic pitch contours in all three dialects, varying only in relative average pitch. Of the remaining two, the drop tone alone displays variations in contour within the same pitch range, while the other, the broken tone, occurs as a separate tone only in one dialect.

To test the above hypotheses, I conducted a series of experiments involving both real speech and synthetic tones. The purpose of the experiments was not to replicate similar perceptual experiments in other languages, but to test tone perception in a rather different situation, by exposing native subjects to the tones of all three dialects in successive tests.

The four experiments described in the next section were designed to test the perception of tones in different ways. Experiment I tested the perception of real speech tones in meaningful context, i.e. tones at the lexical level, when lexical meaning can be inferred from meaningful utterances and contribute to the identification of tones. It was predicted that recognition scores for this test would be highest, although differences may exist between dialects. Experiment II tested the perception of real speech tones in isolated syllables, where the phonetic features of the tones alone would determine their recognition. It was predicted that recognition scores would be lower than in Experiment I, and subjects would do best in identifying tones of their own dialects. Experiment III tested the perception of synthetic tones with Fo variations only, where other clues were excluded. It was predicted that overall recognition scores would be lowest, although for some tones, it was expected that Fo alone would be sufficient for their identification. Experiment IV tested the perception of synthetic tones with Fo and I (intensity) variations, and it was expected that recognition scores would be improved as compared with Experiment III.

This chapter will present and discuss the results of these experiments. Details of the experiments follow in the next sections.

4.1 PROCEDURES

The four experiments described below differed mainly in their contents. The common procedures are described in detail in 4.1.5 and 4.1.6.

4.1.1 Materials for Experiment I

(Real Speech Tones in Context)

Three different lists, one for each of the dialects, were prepared for this experiment. They included two- to six-syllable utterances of meaningful words, phrases or sentences, mainly of standard Vietnamese, with 10% of syllable-morphemes characteristic of dialectal usage in each list. The actual lists and the occurrences of different tones in each list are given in Tables 4.1 and 4.2 (following pages).

The lists were recorded by a male native speaker for each dialect, namely by Informants NM1, CM2 and SM2. The informants were instructed to record the list at dictation speed with a short pause after each utterance, which resulted in an average rate of about sixty syllables per minute, including pauses.

4.1.2 Materials for Experiment II

(Real Speech Tones in Isolated Syllables)

The syllable /ta/, occurring with all non-stopped tones in the following words

/tā/	/tà/	/tá/	/tạ/	/tả/	/tã/
'we'	'oblique'	'dozen'	'quintal'	'describe'	'nappy'

was used for the tests. The stopped variants of the rising and drop tones were excluded because they don't occur in the same segmental environment as the above syllable. Three different lists,

TABLE 4.1 Test-Lists of NV, CV and SV Utterances (in Official Orthography) for Experiment I (Real Speech Tones in Context)

(a) NV

tồi tệ	lờ lững	tĩnh tại	lửng lơ	nhí nhảnh
'wretched'	'adrift'	'stationary'	'overhanging'	'cute'
hai gò má	có đôi mã	nhìn mà xem		
'both cheeks'	'have two horses'	'look and see'		
chẳng phải ma	trong mồ mả	trên ruộng mạ		
'it's not a ghost'	'in the graves'	'on the rice seed-bed'		
đi lại nhộn nhịp	gánh gồng kĩu kịt	vó câu khấp khểnh		
'come and go busily'	'carried on squeaky poles'	'uneven horse trots'		
bánh xe gập ghềnh	kỷ luật chặt chẽ	cải cách triệt để		
'wheels jolting along'	'strict discipline'	'thoroughgoing reform'		
trong đống sách vở cũ	có nhiều tài liệu cần giữ			
'in the stack of old books'	'there are many documents to be preserved'			
chỗ nào chưa hiểu rõ	xin cứ hỏi lại cụ thể			
'where not clearly understood'	'please ask again specifically'			

(b) CV

tả tơi	cũ kỹ	tỉnh táo	tàn tạ	lặng lẽ
'in tatters'	'old-looking'	'awake'	'decaying'	'silent'
ở xứ lạ	uống nước lã	hút thuốc lá		
'in foreign lands'	'drink pure water'	'smoke cigarettes'		
người đói lả	không la lối	thật là lạ		
'be famishing'	'not clamouring'	'it's really strange'		
tiếng tăm lững lẫy	tự do tư tưởng	tương tự như nhau		
'great reputation'	'free thinking'	'similar to each other'		
tướng tá cao cấp	tưởng tượng hão huyền	tỏ tường ngành ngọn		
'high-ranking officers'	'wild imagination'	'knowing thoroughly'		
mấy bữa ni ở nhà	mạ tôi không về bến nớ			
'these few days at home'	'my mother did not go back there'			
hắn mới tới bữa tê	không có người mô mở cửa			
'he just arrived the other day'	'there is nobody to open the door'			

(continued next page)

TABLE 4.1 (continued)

(c) SV

cây cối	cằn cỗi	lạnh lẽo	lảo đảo	kỹ lưỡng
'vegetation'	'barren'	'chilly'	'stagger'	'carefully'

buổi chiều tà	mưa tầm tã	thật khó tả
'declining evening'	'rain cats and dogs'	'really difficult to describe'
chị em ta	mua nửa tá	bán sáu tạ
'we sisters'	'buy half a dozen'	'sell six quintals'
án nói lung tung	thái độ lúng túng	đồ đạc lủng củng
'wanton talk'	'embarrassed attitude'	'things in disorder'
bước đi lững thững	áo quần lụng thụng	mặt mũi bơ phờ
'leisurely gait'	'baggy clothes'	'haggard looks'

cậu nói chẳng rõ ràng	tôi đâu có hiểu được hết
'you don't speak clearly'	'I cannot understand all'
mai mốt hẹn chú sang	có chi cùng nhậu nhẹt hè
'next time I ask you to come over'	'(we'll) have a drink together'

TABLE 4.2 Occurrences of Tones in Test-Lists of Experiment I

TONES		DIALECTS NV	CV		SV	
Level	/-/	13	15		13	
Falling	/\`/	12	11		13	
Rising*	/ʼ/	12	19		12	
Drop*	/./	15	10		15	
Curve**	/ˀ/	13	11	19	9	18
Broken**	/~/	9	8		9	
		74	74		74	

* These tones include both the basic (sonorant-ending) variants and stopped variants.

** These forms are phonetically the same, in CV and SV, but marked with different diacritics in the test-lists according to standard spelling.

one for each dialect , were prepared for this experiment. Each list included 60 syllables in randomized order (6 tones x 10 occurrences each for NV, 5 tones x 12 occurrences each for CV and SV.) They were grouped in series of five, so that they could be recorded with a pause after each series. They were recorded by one male and one female informant for each dialect, namely by Informants NM1, NF1, CM2, CF4, SM2 and SF5, at about the same speed as for Experiment I.

The Fo data of the tones, as produced by these informants, were obtained from measurements of three tokens of each tone. They are plotted in Figures 4.1, 4.2 and 4.3 (pp 194-196), in section 4.3 where the results are discussed.

4.1.3 Materials for Experiment III

(Synthetic Tones With Fo Variations Only)

For the purpose of this experiment, synthetic speech signals approximating the Vietnamese syllable /ta/ were produced by a computer-synthetizer at the Australian National University Computer Centre. The hardware consisted mainly of a TEKTRONIX 4014 graphics terminal connected on the one hand to a VERSATEC 1200A plotter via a VERSATEC C-TEX-2 interface, and on the other hand to a PDP-11/40 computer system which is in turn connected to a REVOX A17 tape recorder via an ADAC-16 signal interface and a ROCKLAND analogue filter. The software consisted of four programs, TABDYN for parameter input via tablet, SYNTH for synthetizing the input into speech signals, RASP for randomizing the sequences and WAVE for audio-visual waveform editing (Millar 1974, 1978).

Formant frequencies derived from spectrograms of the level tone syllable /ta/ by the three male informants, NM1, CM2 and SM2, formed the basis for the synthetic syllables, on which Fo curves approximating those produced by the same informants were superimposed.

Although I was aware that there is interaction between tone and vowel quality (Lehiste 1970, Zee 1978), and this was apparent in the formant shifts in my own spectrograms (see Ch. Two), I decided that for the purpose of my experiments this might be ignored. I therefore used the same formant structures for the syllable /ta/ with the NV level tone for synthetizing syllables with all other tones. Interestingly enough, spectrograms made from the synthetic tones show the formant shifts similar to those of natural tones (see Figure 4.9, p 208 infra). This suggests that some form of interaction between Fo and formant frequencies occurs during the synthetizing process.

I also set the average duration for each syllable at 30 centiseconds. The intensity curve was that of the level tone, which was similar in all three dialects. As there was no tested way of simulating laryngealization in synthetic speech, I decided to create the impression of creaky voice by synthetizing a sharp drop in Fo in the middle of the NV broken tone (as it happened with Fo curves of some NV informants.) For the creaky ending of the NV drop tone and the CV curve tone, the drop in Fo was followed by wavy Fo lines, simulating the irregular Fo drops and jumps in some informants' mingograms.

Three recordings, one for each dialect, were made for the tests, each including 60 synthetized syllables in randomized order (six tones x 10 occurrences each for NV, 5 tones x 12 occurrences each for CV and SV.) The speech signals were recorded at the rate of one syllable per second with a pause of five seconds after each series of five. This resulted in fact in a somewhat slower rate than for Experiments I and II.

4.1.4 *Materials for Experiment IV*
(Synthetic Tones With Fo and Intensity Variations)

For the purpose of this experiment, the same synthetic syllable /ta/ with different NV, CV and SV Fo curves used in Experiment III was modified by varying the intensity parameter I. Data concerning the I contour of individual tones were derived from measurements of mingograms of intensity curves described in Ch. Two. They are shown in normalized forms in Figure 4.6 (p 201). As input parameters to be synthetized, they were given dB values between 0 and 100 corresponding to percentages in Figure 4.6. Note that the creaky portion of the NV drop tone and the CV curve tone was drawn again as wavy lines on the I curves to correspond to the wavy lines on the Fo curves, and the midpoint break of the NV broken tone was represented as a drop in intensity of 30 dB from a peak of 70 dB, then a rise of 20 dB to 60 dB (within a range of 100 dB provided by the computer program.)

These synthetized speech signals were recorded in the same way as for Experiment III to make three tests, one for each dialect.

4.1.5 *Testing Conditions and Subjects*

After pilot tests had been conducted with some subjects in Canberra, minor adjustments were made in the test materials. A pause of five seconds was inserted after every series of five syllables in all experiments except Experiment I where it was inserted after each meaningful utterance. Sample signals of the synthetic tones were added to Experiment III and IV before the actual test signals. The experiments were then conducted in three places : Hanoi, Hue and Ho Chi Minh City, with NV, CV and SV subjects respectively.

In Hanoi and Ho Chi Minh City, the tests were administered in Tandberg language laboratories where subjects were equipped with headphones. In Hue, the tests were conducted in ordinary classrooms and the tapes were played from a portable UHER 4000 Reporter IC Recorder. To enhance listening conditions in such a situation, smaller groups of subjects were tested at a time, sitting near to the recorder.

Subjects included 26 NV, 30 CV and 28 SV speakers (some of whom did not attend the last session for Experiment IV.) All subjects were university students mostly in their late teens and early twenties. Although they came from different parts of the country and represented all major population areas, they were speakers of the standard dialects and were fairly well exposed to the other dialects through the media, education or travel.

4.1.6 Test Procedures

Before each session, I explained to the subjects what was expected from them in the experiments. I told them beforehand that they should expect to hear NV, CV and SV tones in different tests, but did not tell them of the order of presentation, nor that synthetic speech was involved. The idea was that, first, they had to work out for themselves which dialects they were hearing and try to identify the tones from their phonetic and phonological features. Secondly I thought that the quality of the synthetic tones was good and if they were not warned of this beforehand, they would probably ignore any minor peculiarities they migh detect. (Actually, none of them had heard of synthetic speech before, and only a few told me afterwards that they detected some peculiar speech quality in parts of the tests.)

The same order was however presented to the three groups of subjects, i.e. Experiments I, II, III and IV successively, and a different order of NV, CV and SV tones for each experiment. For the synthetic speech tests, a series of sample signals, three for each tone, were given for the subjects to hear before each test-list was played out.

Subjects were issued with test sheets on which the test-lists of syllables in standard Vietnamese spelling were printed without tone marks. They were asked to use the usual Vietnamese diacritics to mark the tones they thought they heard, except that for the level tone, which is left unmarked in Vietnamese spelling, they were to use the macron (so that an unmarked syllable would be later counted as a non-response.)

4.2 RESULTS

The results of the four experiments are given in Tables 4.3 to 4.14 on the following pages.

Tables 4.3, 4.4 and 4.5 give the recognition percentages of real speech tones in context of NV, CV and SV respectively. The first column of each table indicates the group of subjects involved and their number. The second column gives the tone stimuli used in the test in question. The next six columns give the percentage of responses for each tone, with the correct responses underlined. The last but one column gives the actual number of responses for each tone on which the percentages were calculated. The last column gives the total number of responses, that of correct responses and the percentages of correct responses for each group of subjects. Tables 4.6 to 4.8, 4.9 to 4.11 and 4.12 to 4.14 present the results of Experiments II, III and IV respectively in the same way, except that combined results for all three groups of subjects were also given in tabular form at the bottom of each table.

Concerning the definition of correct responses, it should be noted that it was somewhat different according to dialect because of the different number of tones.

For Experiment I, which tested the perception of tone at the lexical level, the differentiation between the curve tone /'/ and the broken tone /~/ was valid for NV tones heard by NV subjects only. Each of them, when marked by either diacritic by CV and SV subjects, was counted as correct because there is no differentiation in their dialects and this would not affect word recognition. On the other hand, CV and SV curve tone /'/ when marked by either diacritic by any subject was also counted as correct for the same reason, and in fact the correct percentages for this tone were lumped together in my tables. The point is that in meaningful context, the use of either diacritic in this case is a matter of spelling habits and not of auditory perception.

TABLE 4.3 Recognition of Real Speech NV Tones (in Meaningful Context) by NV, CV and SV Subjects

Subj.	Stim. (NV)	% Responses as /-/	/\/	/'/	/./	/?/	/~/	n Resp.	Totals
NV (26)	/-/ ˧	100.0						331	R: 1908
	/\/ ˨˩	1.0	99.0					310	Cor R:1874
	/'/ ˧˥			100.0				312	% Cor:98.2
	/./ ˨˩̰	0.5		.	99.5			385	
	/?/ ˨˩˦					95.6		338	
	/~/ ˧˥̰			2.6	0.4	3.0	94.0	232	
CV (30)	/-/ ˧	98.7	1.0				0.3	380	R: 2189
	/\/ ˨˩	1.4	97.7	0.3	0.3	0.3		356	Cor R:2148
	/'/ ˧˥			100.0				355	% Cor:98.1
	/./ ˨˩̰	0.7	0.5	0.2	98.4		0.2	446	
	/?/ ˨˩˦ + /~/ ˧˥̰	0.2	0.1	2.8	0.1	96.8		652	
SV (28)	/-/ ˧	99.1	0.9					345	R: 1961
	/\/ ˨˩	2.2	97.5		0.3			313	Cor R:1925
	/'/ ˧˥			99.7			0.3	325	% Cor:98.1
	/./ ˨˩̰	0.5	0.5	0.3	98.7			388	
	/?/ ˨˩˦ + /~/ ˧˥̰	0.5		1.5	1.2	96.8		590	

For all 84 subjects

Total responses: 6058

Correct responses: 5947

% correct 98.2

TABLE 4.4 Recognition of Real Speech CV Tones (in Meaningful Context) by NV, CV and SV Subjects

Subj.	Stim. (CV)	% Responses as						n Resp.	Totals
		/-/	/`/	/´/	/./	/ʼ/	/~/		
NV	/-/ ˥	98.4	0.7	0.5	0.2	0.2		429	R: 1862
(26)	/`/ ˧	1.2	93.6	1.2	1.2	2.8		250	Cor R:1735
	/´/ ˧˥	0.8	1.0	96.5	0.7	1.0		484	% Cor:93.1
	/./ ˩	1.2	11.3	0.4	80.6	6.5		248	
	/ʼ/ ˩˧	2.9	2.2	0.9	2.7	91.3		451	
CV	/-/ ˥	99.6		0.4				493	R: 2136
(30)	/`/ ˧	1.0	96.8	0.4	1.0	0.8		284	Cor R:2026
	/´/ ˧˥	1.3	2.5	94.6	0.6	1.0		556	% Cor:94.8
	/./ ˩	1.0	2.8	0.4	93.0	2.8		286	
	/ʼ/ ˩˧	1.0	1.9	0.4	6.2	90.5		517	
SV	/-/ ˥	98.8	0.2	0.8		0.2		424	R: 1816
(28)	/`/ ˧	0.4	98.3	0.9		0.4		238	Cor R:1737
	/´/ ˧˥	1.0	0.7	95.1	1.5	1.7		468	% Cor:95.6
	/./ ˩		2.4	1.6	95.5	0.5		246	
	/ʼ/ ˩˧	1.6	0.3		6.3	91.8		440	

For all 84 subjects

Total responses: 5814

Correct responses: 5498

% correct 94.5

TABLE 4.5 Recognition of Real Speech SV Tones (in Meaningful Context) by NV, CV and SV Subjects

Subj.	Stim. (SV)	% Responses as /-/	/\/	/'/	/./	/ʼ/	/~/	n Resp.	Totals
NV (26)	/-/ ˧	100.0						386	R: 1920
	/\/ ˨˩	0.4	97.7	0.4		1.5		260	Cor R:1887
	/'/ ˧˥	1.1		97.7	0.9	0.3		442	% Cor:98.3
	/./ ˨˩˨	3.3		0.3	96.4			364	
	/ʼ/ ˨˩˦	0.2		0.2	0.4	99.2		468	
CV (30)	/-/ ˧	100.0						440	R: 2187
	/\/ ˨˩		97.6		0.4	2.0		297	Cor R:2146
	/'/ ˧˥	0.6		98.4	0.6	0.4		503	% Cor:98.1
	/./ ˨˩˨	4.3		0.2	94.5	1.0		416	
	/ʼ/ ˨˩˦			0.4	0.8	98.8		531	
SV (28)	/-/ ˧	99.8		0.2				413	R: 2008
	/\/ ˨˩		97.8	0.4	0.4	1.4		269	Cor R:1988
	/'/ ˧˥	0.7		99.1	0.2			460	% Cor:99.0
	/./ ˨˩˨	1.0			98.4	0.6		375	
	/ʼ/ ˨˩˦	0.6				99.4		491	

For all 84 subjects

Total responses: 6115

Correct responses: 6030

% correct: 98.6

TABLE 4.6 Recognition of Real Speech NV Tones (Isolated Syllables) by NV, CV and SV Subjects

Subj.	Stim. (NV)	% Responses as						n Resp.	Totals
		/-/	/\/	/´/	/./	/ʾ/	/~/		
NV (26)	/-/ ˧	99.4	0.2		0.4			515	R: 3100
	/\/ ˨˩	0.8	98.0	0.2	0.8		0.2	517	Cor R:2865
	/´/ ˧˥	0.2		88.4	0.2	2.3	8.9	517	% Cor:92.4
	/./ ˨˩ˀ		0.2	0.2	94.6	0.4	4.6	519	
	/ʾ/ ˧˩˧		8.1	0.2	6.2	82.0	3.5	517	
	/~/ ˧ˀ˥	0.2		1.6	0.2	6.0	92.0	515	
CV (30)	/-/ ˧	97.6	1.2	0.5	0.2	0.3	0.2	589	R: 3542
	/\/ ˨˩	4.2	86.4	0.5	7.9	0.8	0.2	596	Cor R:2713
	/´/ ˧˥	0.5	1.2	89.0	0.5	5.4	3.4	590	% Cor:76.6
	/./ ˨˩ˀ	1.2	3.5	0.2	80.5	7.8	6.8	589	
	/ʾ/ ˧˩˧	0.7	5.6	1.1	23.3	51.3	18.0	589	
	/~/ ˧ˀ˥	0.8	0.5	2.2	1.5	40.3	54.7	589	
SV (28)	/-/ ˧	99.0	0.2	0.8				534	R: 3211
	/\/ ˨˩	2.2	93.4	0.2	3.8	0.4		531	Cor R:2584
	/´/ ˧˥	1.2	0.5	85.8		5.9	6.6	542	% Cor:80.3
	/./ ˨˩ˀ		0.5		93.8	3.9	1.8	544	
	/ʾ/ ˧˩˧	0.2	9.8		22.4	51.6	16.0	531	
	/~/ ˧ˀ˥	0.4		3.2	0.4	37.6	58.4	531	
All V (84)	/-/ ˧	98.8	0.5	0.4	0.2	0.1		1638	R: 9853
	/\/ ˨˩	2.5	92.4	0.3	4.3	0.4	0.1	1644	Cor R:8162
	/´/ ˧˥	0.6	0.6	87.8	0.2	4.6	6.2	1649	% Cor:82.8
	/./ ˨˩ˀ	0.4	1.5	0.1	89.3	4.2	4.5	1652	
	/ʾ/ ˧˩˧	0.3	7.8	0.4	17.6	61.2	12.7	1637	
	/~/ ˧ˀ˥	0.5	0.2	2.3	0.7	28.6	67.7	1635	

TABLE 4.7 Recognition of Real Speech CV Tones (Isolated Syllables) by NV, CV and SV Subjects

Subj.	Stim. (CV)	% Responses as /-/	/\/	/'/	/./	/ʼ/	/~/	n Resp.	Totals
NV (26)	/-/ ˥	96.0	0.6	2.2	0.5	0.5	0.2	620	R: 2993
	/\/ ˧	6.0	92.3		0.8	0.3	0.6	601	Cor R:1968
	/'/ ˧˥	1.0	0.2	25.0	1.0	38.1	34.7	599	% Cor:65.8
	/./ ˨˩	1.8	57.8	0.3	36.1	3.5	0.5	571	
	/ʼ/ ˨˩˦	0.8	5.0	1.2	16.3	22.1	54.6	602	
						76.7			
CV (30)	/-/ ˥	92.3	2.0	2.3	1.6	1.1	0.7	693	R: 3417
	/\/ ˧	3.3	78.6	1.6	10.5	4.8	1.2	696	Cor R:2483
	/'/ ˧˥	3.5	1.6	53.6	4.0	20.7	16.6	677	% Cor:72.7
	/./ ˨˩	2.9	26.3	1.8	57.6	9.0	2.4	658	
	/ʼ/ ˨˩˦	3.2	4.6	3.3	9.0	53.2	26.7	693	
						79.9			
SV (28)	/-/ ˥	93.1	1.1	4.4	0.6	0.8		635	R: 3115
	/\/ ˧	8.1	80.5	0.3	8.5	2.1	0.5	626	Cor R:1882
	/'/ ˧˥	2.4	1.3	23.3	1.8	44.3	26.9	621	% Cor:60.4
	/./ ˨˩	3.5	64.9	0.3	26.2	3.9	1.2	596	
	/ʼ/ ˨˩˦	2.8	7.2	3.5	10.2	42.7	33.6	617	
						76.3			
All V (84)	/-/ ˥	93.7	1.3	3.0	0.9	0.8	0.3	1948	R: 9525
	/\/ ˧	5.7	83.5	0.7	6.8	2.5	0.8	1923	Cor R:6333
	/'/ ˧˥	2.4	1.0	34.7	2.3	33.9	25.7	1877	% Cor:66.5
	/./ ˨˩	2.7	48.8	0.9	40.6	5.6	1.4	1825	
	/ʼ/ ˨˩˦	2.3	5.6	2.7	11.6	40.1	37.7	1932	
						77.8			

TABLE 4.8 Recognition of Real Speech SV Tones (Isolated Syllables) by NV, CV and SV Subjects

Subj.	Stim. (SV)	% Responses as /-/	/\/	/ʻ/	/./	/ʼ/	/~/	n Resp.	Totals
NV (26)	/-/ ˧	99.0	0.8		0.2			588	R: 3069
	/\/ ˨˩	2.2	89.1	0.3	5.4	2.8	0.2	633	Cor R:2585
	/ʻ/ ˧˥	4.0	0.3	88.0	0.5	0.3	6.9	609	% Cor:84.2
	/./ ˨˩˨	0.3	6.4	0.8	72.3	16.8	3.4	611	
	/ʼ/ ˨˩˦	1.4	1.8	20.4	3.0	47.8	25.6	628	
						73.4			
CV (30)	/-/ ˧	95.1	1.8	1.1	1.1	0.6	0.3	656	R: 3449
	/\/ ˨˩	2.0	65.0	1.0	25.1	5.9	1.0	693	Cor R:2562
	/ʻ/ ˧˥	8.0	0.5	87.2	1.7	1.6	1.0	690	% Cor:74.3
	/./ ˨˩˨	1.3	7.5	4.4	58.5	22.2	6.1	687	
	/ʼ/ ˨˩˦	1.2	1.7	19.1	11.1	50.3	16.6	723	
						66.9			
SV (28)	/-/ ˧	99.0	0.3	0.7				619	R: 3221
	/\/ ˨˩	0.3	85.1		14.1	0.3	0.2	645	Cor R:2909
	/ʻ/ ˧˥	1.6	0.2	98.0		0.2		652	% Cor:90.3
	/./ ˨˩˨	0.9	2.4	4.0	78.6	10.4	3.7	645	
	/ʼ/ ˨˩˦	0.3	0.6	5.3	4.3	75.6	13.9	660	
						89.5			
All V (84)	/-/ ˧	97.6	1.0	0.7	0.4	0.2	0.1	1883	R: 9739
	/\/ ˨˩	1.5	79.3	0.5	15.2	3.0	0.5	1971	Cor R:8046
	/ʻ/ ˧˥	4.6	0.3	91.1	0.8	0.7	2.5	1951	% Cor:82.6
	/./ ˨˩˨	0.9	5.4	3.1	69.5	15.6	4.5	1943	
	/ʼ/ ˨˩˦	1.0	1.3	15.0	6.3	57.8	18.5	2011	
						76.4			

TABLE 4.9 Recognition of Synthetic NV Tones (With F_0 Variations Only) by NV, CV and SV Subjects

Subj.	Stim. (NV)	% Responses as /-/	/\/	/´/	/./	/ʼ/	/~/	n Resp.	Totals
NV	/-/	97.3			2.7			260	R: 1560
(26)	/\/		38.8		61.2			260	Cor R:1205
	/´/			93.5			6.5	260	% Cor:77.2
	/./		30.8		58.8	10.0	0.4	260	
	/ʼ/		2.7	0.4	7.3	88.5	1.1	260	
	/~/		0.4	7.7		5.4	86.5	260	
CV	/-/	100.0						299	R: 1975
(30)	/\/	0.3	31.2		67.1	0.7	0.7	298	Cor R:1128
	/´/	0.3		89.7		6.7	3.3	300	% Cor:62.8
	/./		15.8	0.3	55.1	23.8	5.0	298	
	/ʼ/		0.6		41.7	48.0	9.7	300	
	/~/			13.7		33.3	53.0	300	
SV	/-/	99.2	0.4		0.4			275	R: 1642
(28)	/\/		53.1		45.8	1.1		271	Cor R:1129
	/´/			87.8	0.4	0.8	11.0	272	% Cor:68.7
	/./	0.8	29.8	0.8	46.5	20.0	2.1	275	
	/ʼ/	1.1	2.1	0.8	19.5	61.7	14.8	277	
	/~/	0.4	0.4	14.2	0.4	20.6	64.0	272	
All V	/-/	98.9	0.1		1.0			834	R: 5005
(84)	/\/	0.1	40.8		58.3	0.6	6.2	829	Cor R:3462
	/´/	0.1		90.3	0.1	2.6	6.9	832	% Cor:69.2
	/./	0.2	25.2	0.4	53.4	18.2	2.6	833	
	/ʼ/	0.4	1.8	0.4	23.6	65.1	8.7	837	
	/~/	0.1	0.2	12.0	0.1	20.2	66.4	840	

TABLE 4.10 Recognition of Synthetic CV Tones (With F_0 Variations Only) by NV, CV and SV Subjects

Subj.	Stim. (CV)	% Responses as /-/	/\/	/ʻ/	/./	/ʼ/	/~/	n Resp.	Totals
NV (26)	/-/ ˥	99.7	0.3					312	R: 1558
	/\/ ˧		35.0		57.7	5.1	1.2	312	Cor R: 947
	/ʻ/ ˧˥			80.7	2.0	2.8	14.5	311	% Cor: 60.8
	/./ ˨˩		0.3		64.1	32.4	3.2	312	
	/ʼ/ ˨˩~		1.2	2.0	73.3	18.3	5.2	311	
						23.5			
CV (30)	/-/ ˥	100.0						359	R: 1789
	/\/ ˧		27.6	0.3	60.3	9.3	2.5	355	Cor R: 1019
	/ʻ/ ˧˥			85.3	0.3	6.9	7.5	360	% Cor: 56.9
	/./ ˨˩	0.3		0.3	35.7	49.7	14.0	356	
	/ʼ/ ˨˩~	0.6	2.5	7.0	54.3	27.3	8.3	359	
						35.6			
SV (28)	/-/ ˥	98.5	0.6	0.6	0.3			333	R: 1654
	/\/ ˧	2.7	67.9		24.8	2.7	1.9	330	Cor R: 981
	/ʻ/ ˧˥			39.4	0.3	30.9	29.4	327	% Cor: 59.3
	/./ ˨˩		0.9	0.6	77.0	10.0	11.5	331	
	/ʼ/ ˨˩~		0.6	5.1	80.8	8.7	4.8	333	
						13.5			
All V (84)	/-/ ˥	99.4	0.3	0.2	0.1			1004	R: 5001
	/\/ ˧	0.9	43.5	0.1	47.8	5.8	1.9	997	Cor R: 2947
	/ʻ/ ˧˥			68.8	0.7	13.6	16.9	998	% Cor: 58.9
	/./ ˨˩	0.1	0.5	0.2	58.3	31.1	9.8	999	
	/ʼ/ ˨˩~	0.2	1.5	4.8	69.0	18.3	6.2	1003	
						24.5			

TABLE 4.11 Recognition of Synthetic SV Tones (With F_o Variations Only) by NV, CV and SV Subjects

Subj.	Stim. (SV)	% Responses as /-/	/\/	/'/	/./	/ʼ/	/~/	n Resp.	Totals
NV (26)	/-/ ˧	97.7			2.3			306	R: 1550
	/\/ ˨˩	0.3	33.8	0.3	65.3	0.3		311	Cor R:1134
	/'/ ˧˥			95.8			4.2	312	% Cor:72.5
	/./ ˨˩˨	0.6	3.6	0.3	67.0	28.2	0.3	309	
	/ʼ/ ˨˩˦			30.8	0.6	36.2	32.4	312	
						68.6			
CV (30)	/-/ ˧	100.0						359	R: 1789
	/\/ ˨˩	0.3	40.9		55.2	2.5	1.1	359	Cor R:1342
	/'/ ˧˥			97.2		2.2	0.6	360	% Cor:75.0
	/./ ˨˩˨		5.1	0.6	62.4	29.4	2.5	354	
	/ʼ/ ˨˩˦		0.9	15.1	9.8	45.4	28.8	357	
						74.2			
SV (28)	/-/ ˧	99.1		0.3	0.6			329	R: 1633
	/\/ ˨˩	0.3	54.4	0.3	44.4	0.3	0.3	331	Cor R:1410
	/'/ ˧˥			97.9	0.3		1.8	325	% Cor:86.4
	/./ ˨˩˨	0.3	0.3	0.6	80.7	13.1	5.0	321	
	/ʼ/ ˨˩˦	0.3	0.9	7.6	3.4	56.3	31.5	327	
						87.8			
All V (84)	/-/ ˧	98.9		0.1	1.0			994	R: 4972
	/\/ ˨˩	0.3	43.2	0.2	54.7	1.1	0.5	1001	Cor R:3835
	/'/ ˧˥			97.0	0.1	0.8	2.1	997	% Cor:77.1
	/./ ˨˩˨	0.3	3.1	0.5	69.8	23.7	2.6	984	
	/ʼ/ ˨˩˦	0.1	0.6	17.6	4.9	46.1	30.7	996	
						76.8			

TABLE 4.12 Recognition of Synthetic NV Tones (With F_0 and Intensity Variations) by NV, CV and SV Subjects

Subj.	Stim. (NV)	% Responses as /-/	/\/	/'/	/./	/ʾ/	/~/	n Resp.	Totals
NV (26)	/-/ ˧	92.7			7.3			259	R: 1558
	/\/ ˨˩		32.3	0.4	66.5	0.4	0.4	260	Cor R:1378
	/'/ ˧˥			82.2		1.6	16.2	259	% Cor:82.0
	/./ ˨	0.4	3.4		95.8	0.4		260	
	/ʾ/ ˧˩		0.4		0.4	96.1	3.1	260	
	/~/ ˧˥		0.4	0.8	0.8	5.0	93.0	260	
CV (30)	/-/ ˧	98.6	0.7		0.7			297	R: 1786
	/\/ ˨˩		27.7	0.3	70.0	1.7	0.3	296	Cor R:1219
	/'/ ˧˥			80.8	0.3	7.8	11.1	297	% Cor:68.2
	/./ ˨		9.7		81.9	8.4		299	
	/ʾ/ ˧˩	0.3	1.1	0.3	32.4	55.5	10.4	299	
	/~/ ˧˥			3.3	1.4	30.5	64.8	298	
SV (22)	/-/ ˧	98.5		0.5	0.5	0.5		220	R: 1316
	/\/ ˨˩		27.0		73.0			219	Cor R: 986
	/'/ ˧˥			86.8	0.5	0.5	12.2	220	% Cor:74.9
	/./ ˨		8.6	0.5	88.6	1.8	0.5	219	
	/ʾ/ ˧˩			1.8	11.9	70.8	15.5	219	
	/~/ ˧˥			1.4		21.0	77.6	219	
All V (78)	/-/ ˧	96.7	0.3	0.1	2.8	0.1		776	R: 4660
	/\/ ˨˩		29.0	0.3	69.6	0.8	0.3	775	Cor R:3483
	/'/ ˧˥			83.0	0.3	3.6	13.1	776	% Cor:74.7
	/./ ˨	0.1	7.3	0.1	88.5	3.9	0.1	778	
	/ʾ/ ˧˩	0.1	0.5	0.7	15.9	73.4	9.4	778	
	/~/ ˧˥		0.1	1.9	0.8	19.3	77.9	777	

TABLE 4.13 Recognition of Synthetic CV Tones (With F_o and **Intensity** Variations) by NV, CV and SV Subjects

Subj.	Stim. (CV)	% Responses as /-/	/\/	/'/	/./	/'/	/~/	n Resp.	Totals
NV (26)	/-/ ˥	100.0						312	R: 1549
	/\/ ˦	0.3	37.4		59.7	2.6		308	Cor R: 993
	/'/ ˧˥	0.3		82.0	0.3	6.8	10.6	311	% Cor:60.2
	/./ ˩		1.0		41.4	55.4	2.2	312	
	/'/ ˩˦		3.3	2.9	53.9	27.7	12.1	306	
						39.9			
CV (30)	/-/ ˥	98.8	0.6		0.3	0.3		359	R: 1789
	/\/ ˦	0.3	35.7	0.6	52.1	9.3	2.0	353	Cor R:1058
	/'/ ˧˥	0.3		85.2		10.3	4.2	359	% Cor:59.1
	/./ ˩	0.3	0.3	1.4	40.0	46.9	11.1	360	
	/'/ ˩˦		7.8	8.4	48.3	24.0	11.5	358	
						35.5			
SV (22)	/-/ ˥	100.0						263	R: 1315
	/\/ ˦	0.4	51.5		44.6	2.3	1.2	264	Cor R: 714
	/'/ ˧˥			40.2		29.1	30.7	264	% Cor:51.9
	/./ ˩	0.8	0.4	0.4	67.0	22.6	8.8	261	
	/'/ ˩˦	0.8	3.8	1.6	80.9	7.6	5.3	263	
						12.9			
All V (78)	/-/ ˥	99.6	0.2		0.1		0.1	934	R: 4653
	/\/ ˦	0.3	40.8	0.2	52.5	5.1	1.1	925	Cor R:2705
	/'/ ˧˥	0.2		71.4	0.1	14.5	13.8	934	% Cor:58.1
	/./ ˩	0.3	0.5	0.6	48.1	43.0	7.5	933	
	/'/ ˩˦	0.2	5.3	4.6	57.4	20.6	9.9	927	
						30.5			

TABLE 4.14 Recognition of Synthetic SV Tones (With F_0 and Intensity Variations) by NV, CV and SV Subjects

Subj.	Stim. (SV)	% Responses as /-/	/\/	/ˊ/	/./	/ʼ/	/~/	n Resp.	Totals
NV (26)	/-/ ˧	95.5			4.5			310	R: 1554
	/\/ ˨˩		29.7	0.6	69.4		0.3	310	Cor R:1000
	/ˊ/ ˧˥			93.3			6.7	312	% Cor:64.3
	/./ ˨˩˨	0.3	1.3		12.8	82.4	3.2	312	
	/ʼ/ ˧˩˥			9.0	0.3	42.6	48.1	310	
						90.7			
CV (30)	/-/ ˧	100.0						358	R: 1798
	/\/ ˨˩		39.2		57.5	3.0	0.3	360	Cor R:1328
	/ˊ/ ˧˥			97.2		0.8	2.0	360	% Cor:73.8
	/./ ˨˩˨			0.6	39.4	50.6	9.4	360	
	/ʼ/ ˧˩˥	0.3		5.3	0.8	53.0	40.6	360	
						93.6			
SV (22)	/-/ ˧	99.7		0.3				264	R: 1318
	/\/ ˨˩		47.9		50.9	0.8	0.4	263	Cor R:1088
	/ˊ/ ˧˥	0.4		98.4	0.4		0.8	264	% Cor:82.5
	/./ ˨˩˨	2.3	0.8	0.8	68.0	18.6	9.5	263	
	/ʼ/ ˧˩˥			1.5		64.4	34.1	264	
						98.5			
All V (78)	/-/ ˧	98.4			1.6			932	R: 4670
	/\/ ˨˩		38.5	0.2	59.6	1.4	0.3	933	Cor R:3416
	/ˊ/ ˧˥	0.1		96.3		0.4	3.2	936	% Cor:73.1
	/./ ˨˩˨	0.8	0.6	0.4	38.6	52.2	7.4	935	
	/ʼ/ ˧˩˥	0.1		5.5	0.4	52.8	41.2	934	
						94.0			

For Experiments II, III and IV, I made a different decision, because they tested perception of tone out of context. For NV tones, as they are phonetically different, the responses to /'/ and /~/ by any subjects were counted as correct only if they correctly differentiated them. For CV and SV tones, the same applied as for Experiment I, but separate percentages of responses as /'/ or /~/ were given in the tables as this point will be important for further discussion of possible perceptual cues.

To compare the test scores within an experiment or between different experiments, I made use of a statistical tool called Chi-square test. The χ^2 value, which reflects the deviation of observed values from expected or theoretical values, is calculated according to the statistical formula

$$\chi^2 = \Sigma \frac{(\,|\,o - e\,|\, - 0.5)^2}{e}$$

where o means observed frequencies,

e means expected frequencies, and

0.5 is the 'Yates correction for continuity,' to be subtracted from the absolute value of o - e, whenever χ^2 is calculated from 1 x 2 or 2 x 2 tables (MacDonald 1977).

The frequencies used in the χ^2 calculations were actual scores of correct and incorrect recognition and not percentages.

4.3 DISCUSSION

Referring to the points mentioned in 4.0, I shall discuss the results of Experiments I - IV in their various aspects and implications. The discussion will consist of three parts. The first concerns the overall recognition scores of real speech tones in the first two experiments; the second is related to perception of synthetic tones; and the third is an analysis of tone features involved in systematic misperceptions in Experiment II.

4.3.1 Tone Perception in Cross-Dialect Situation

The overall percentages of correct recognition for each of the three groups of subjects in all experiments indicate that Vietnamese tones in context can be well identified by speakers of different dialects, and tones in isolated syllables can also be identified fairly well, but there are some differences according to dialects, individual tones and subjects.

They show the general tendency for NV and SV tones to be better identified by all subjects than CV tones.

Even in Experiment I, where identification was facilitated by meaningful context, the percentages of correct recognition were slightly higher for NV and SV tones than for CV tones. There was no significant differences in the recognition scores of NV and SV tones, but the difference was found to be significant at the 0.01 level between CV and the other dialects, when χ^2-tests were applied to the overall recognition scores in this experiment.

The results of Experiment II, where tones in isolated syllables were identified from their perceptual features alone without the help of context, deserve more detailed analysis.

Figures 4.1, 4.2 and 4.3, in which the mean Fo of NV, CV and SV tones, by the six informants whose voices were used in Experiment II, were plotted against normalized duration, are presented on the next pages for easy reference to the results.

Figures 4.4 and 4.5 (following pages) illustrate the results of Experiment II. Figure 4.4 not only illustrates the general tendency noted earlier, but shows that (a) NV and SV subjects identified their own tones best (which was predicted), and each other's tones better than CV tones (the scores were all significantly different for each of the three dialects;) and (b) CV subjects identified their own tones better than the other groups did (which was expected,) but still worse than they did NV tones. The scores were significantly higher for NV tones than those for the other two dialects.

Figure 4.5 shows that mis-identifications occurred often with those tones that display more phonetic variations across dialects. They are the curve tone(which is low-falling-rising in NV, mid-falling-rising-creaky in CV and mid-falling-rising in SV), the drop tone (which is low-falling-creaky in NV, low-falling in CV and low-falling-rising in SV), the broken tone (which is mid or high-falling-rising-creaky and occurs only in NV), and the CV rising tone (which is mid-rising in CV with a fall after onset, while it is high-rising in both NV and SV.)

I would suggest that there are several reasons behind these facts. First, larger perceptual distances exist between the tones in NV and SV (wider pitch range, more extreme contours in both dialects, plus creaky voice in NV,) which made them easier to identify

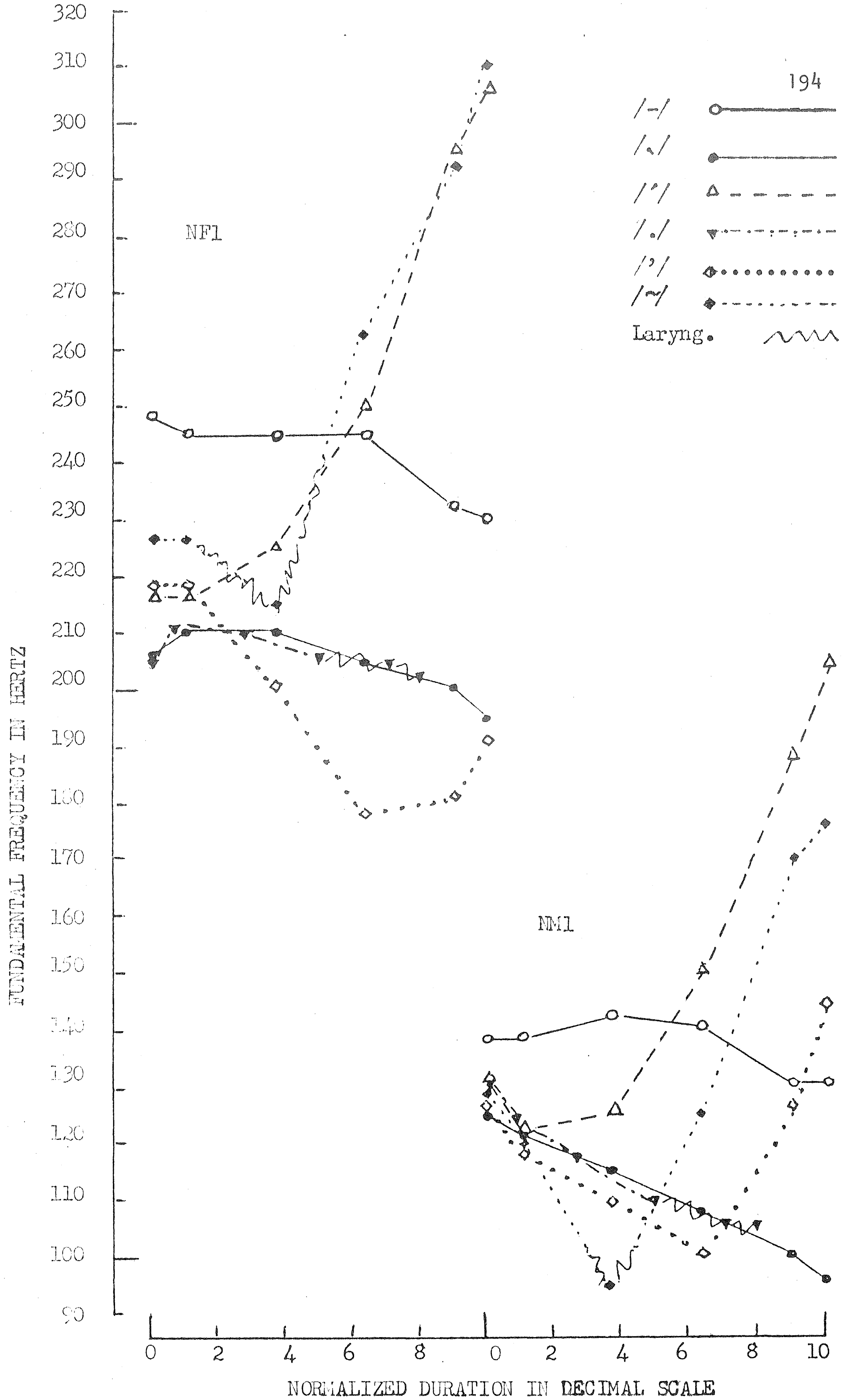


FIGURE 4.1 Mean F_0 of MV Tones by Informants NF1 and NM1 Plotted Against Normalized Duration

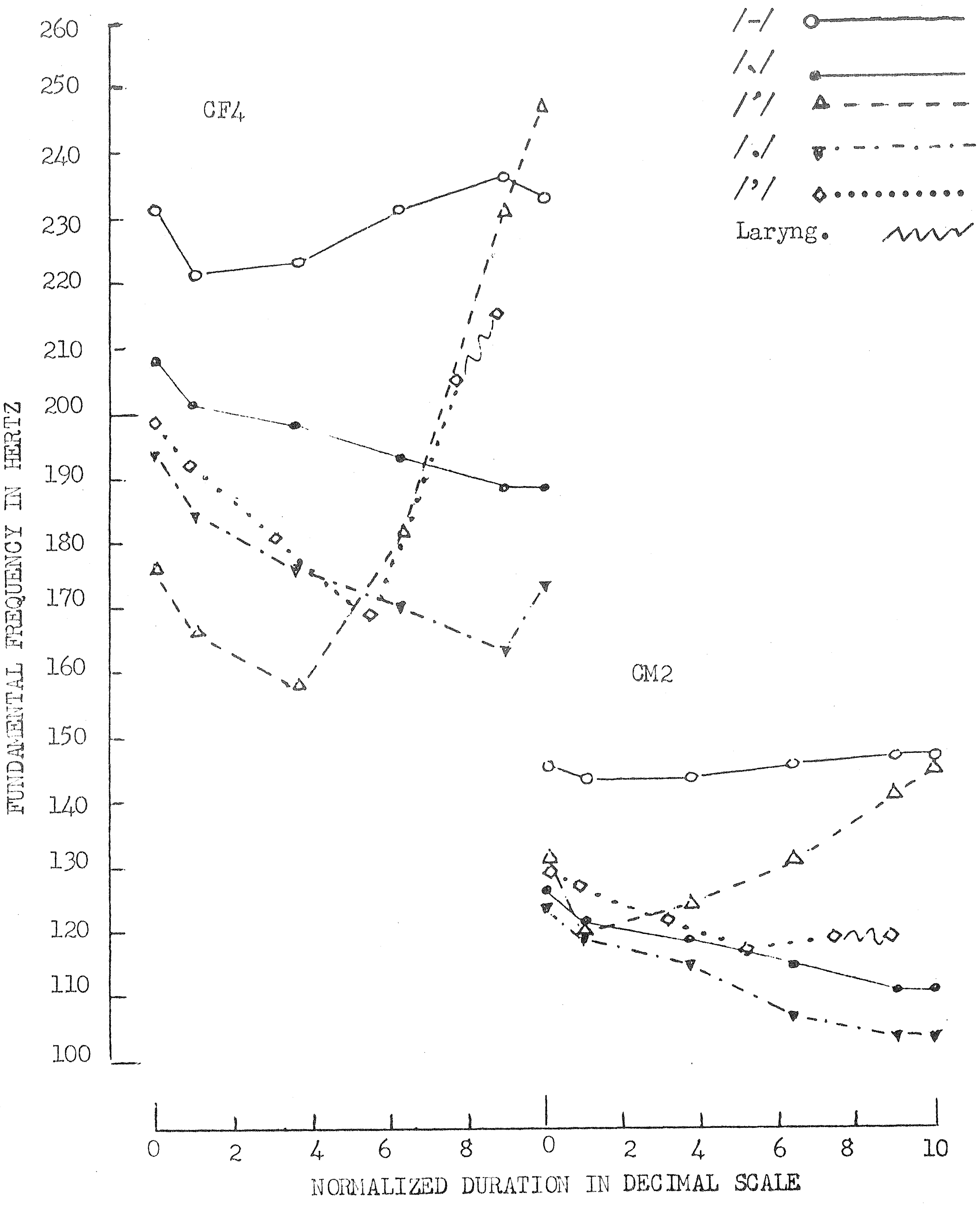


FIGURE 4.2 Mean F_0 of CV Tones by Informants CF4 and CM2 Plotted Against Normalized Duration

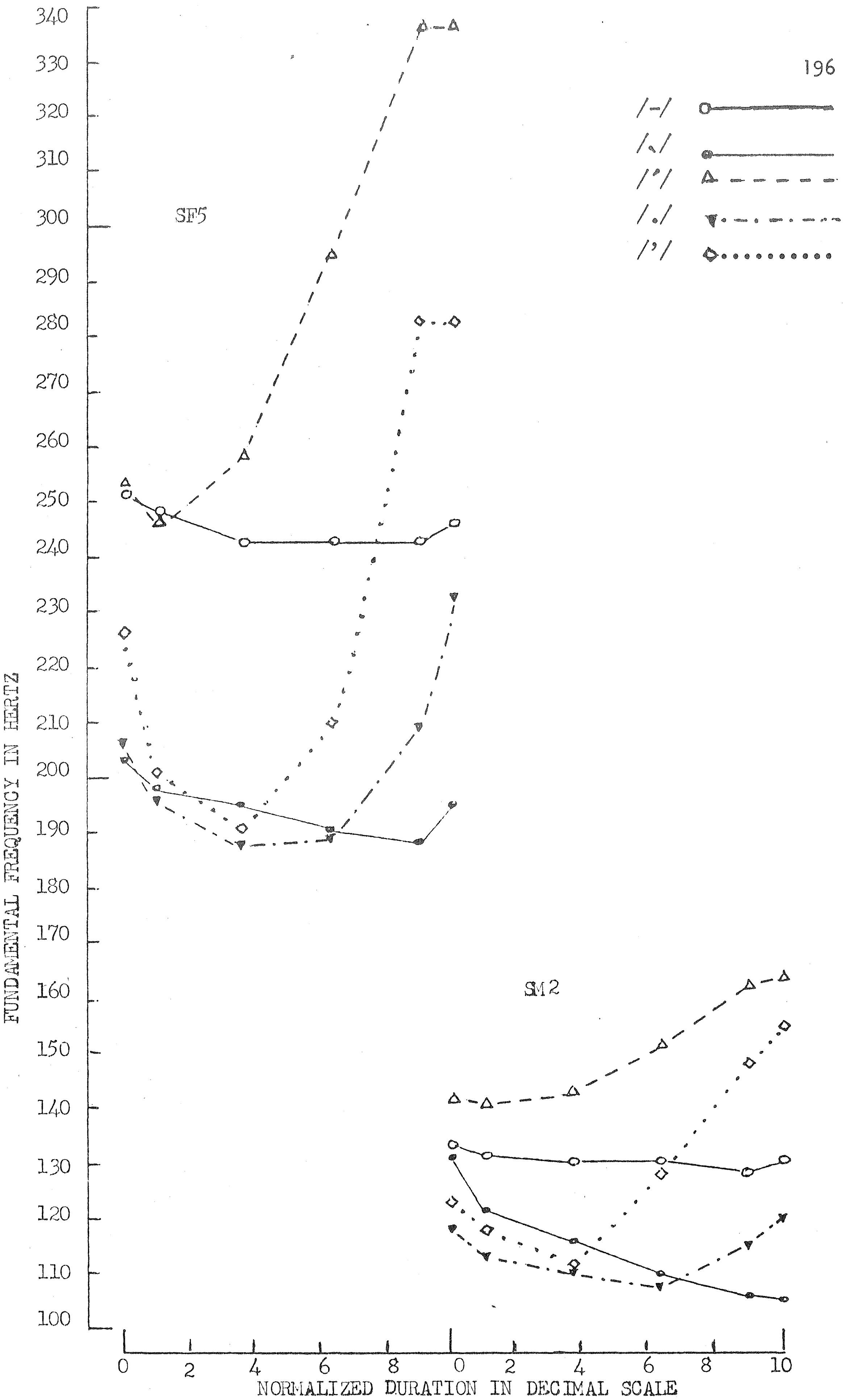


FIGURE 4.3 Mean F_0 of SV Tones by Informants SF5 and SM2 Plotted Against Normalized Duration

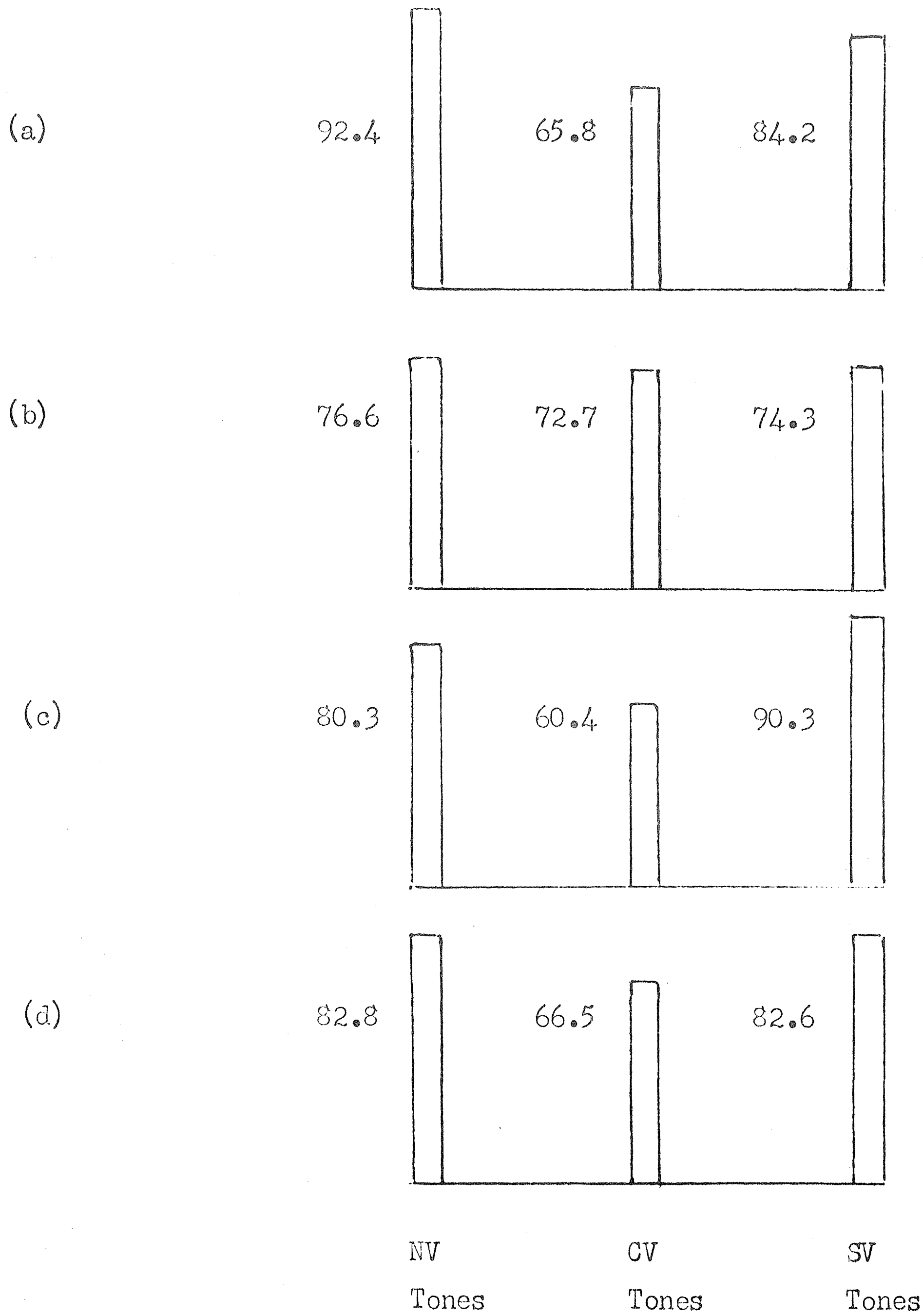


FIGURE 4.4 Overall Percentages of Correct Recognition of Real Speech (Isolated Syllables) Tones by (a) NV, (b) CV, (c) SV and (d) all V Subjects (Data From Tables 4.6, 4.7 and 4.8)

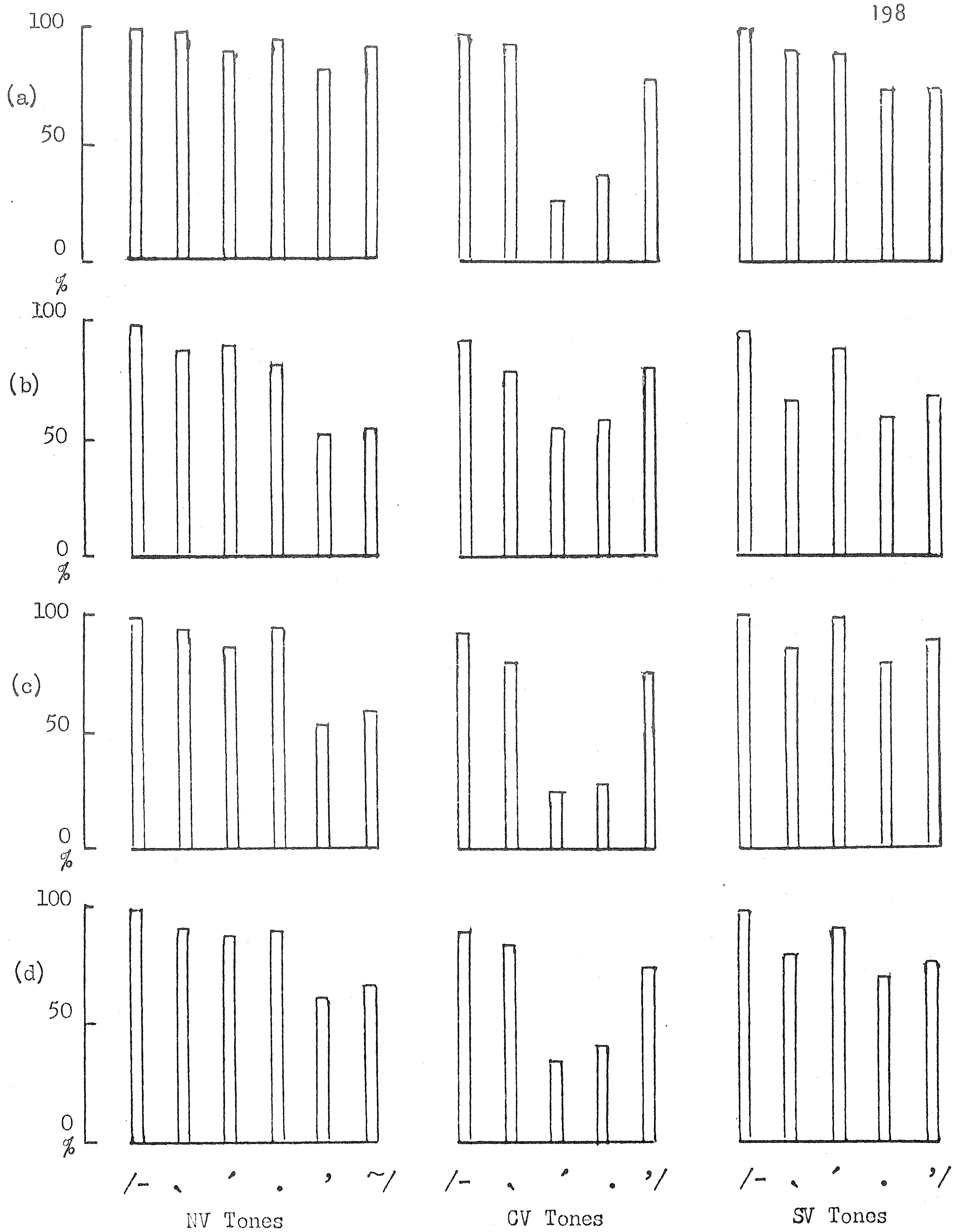


FIGURE 4.5 Percentages of Correct Recognition of Real Speech (Isolated Syllables) Tones by (a) NV, (b) CV, (c) SV and (d) All V Subjects (Data From Tables 4.6, 4.7 and 4.8)

than CV tones (narrower pitch range and predominance of mid-pitch tones.) Second, there are greater similarities between NV and SV tones. Three of them, the level, falling and rising tones, are phonetically similar in both dialects. On the other hand, all CV tones differ from their counterparts in NV and SV either in average pitch level or laryngealization, although most of them share the basic common pitch contours.

These two factors are directly attributable to the acoustic characteristics of the tones as described in Ch. Two (2.3). The third factor is related to the fact that there is more variation between local dialects in CV tones than in NV and SV. Therefore CV subjects might have had more difficulty in identifying the tones of the speakers in the experiments because they may differ from their own in some features (although in my judgment the two speakers chosen, CF4 and CM2, had standard CV tones.) A fourth factor is the special condition of the experiments where the three sets of NV, CV and SV tones were played out for the perceptual tests without warning on the order of presentation. It may be assumed that in such a confusing cross-dialect situation, the more standard and more widely known NV and SV tones were better recognized than the more localized tones of CV. Even the CV subjects themselves might be induced to interpret the tones in terms of the more standard NV and SV forms than their own tones, especially when these forms were given without meaningful context. In this connection, we should bear in mind the fact that NV and SV are each spoken by twenty-odd million people and their areas include Hanoi and Ho Chi Minh City which have been great cultural, educational as well as political and economic centres for the whole country whereas CV is spoken by about five million people only and its areas includes many districts with strong local traditions.

The results also seems to indicate that the subjects' dialectal background had less apparent influence than the other factors. Figure 4.5 shows more similarity in recognition patterns within each tone group than within each subject group. Some evidence for this influence can however be found in details of perceptual errors which show some appreciable difference, such as

NV broken tone /~/ perceived as curve tone /'/

6.0% by NV subjects

40.3% " CV "

37.6% " SV "

CV drop tone /./ perceived as falling tone /\/

57.8% by NV subjects

26.3% " CV "

64.9% " SV "

SV curve tone /'/ perceived as

broken tone /~/ and rising tone /´/ respectively

25.6 and 20.4% by NV subjects

16.6 " 19.1% " CV "

13.9 " 5.3% " SV "

4.3.2 Perception of Synthetic Tones

The results of Experiments III and IV support the general observations for Experiment II.

Figure 4.6 (next page) gives the intensity curves used for individual synthetic tones in Experiment IV, as different from

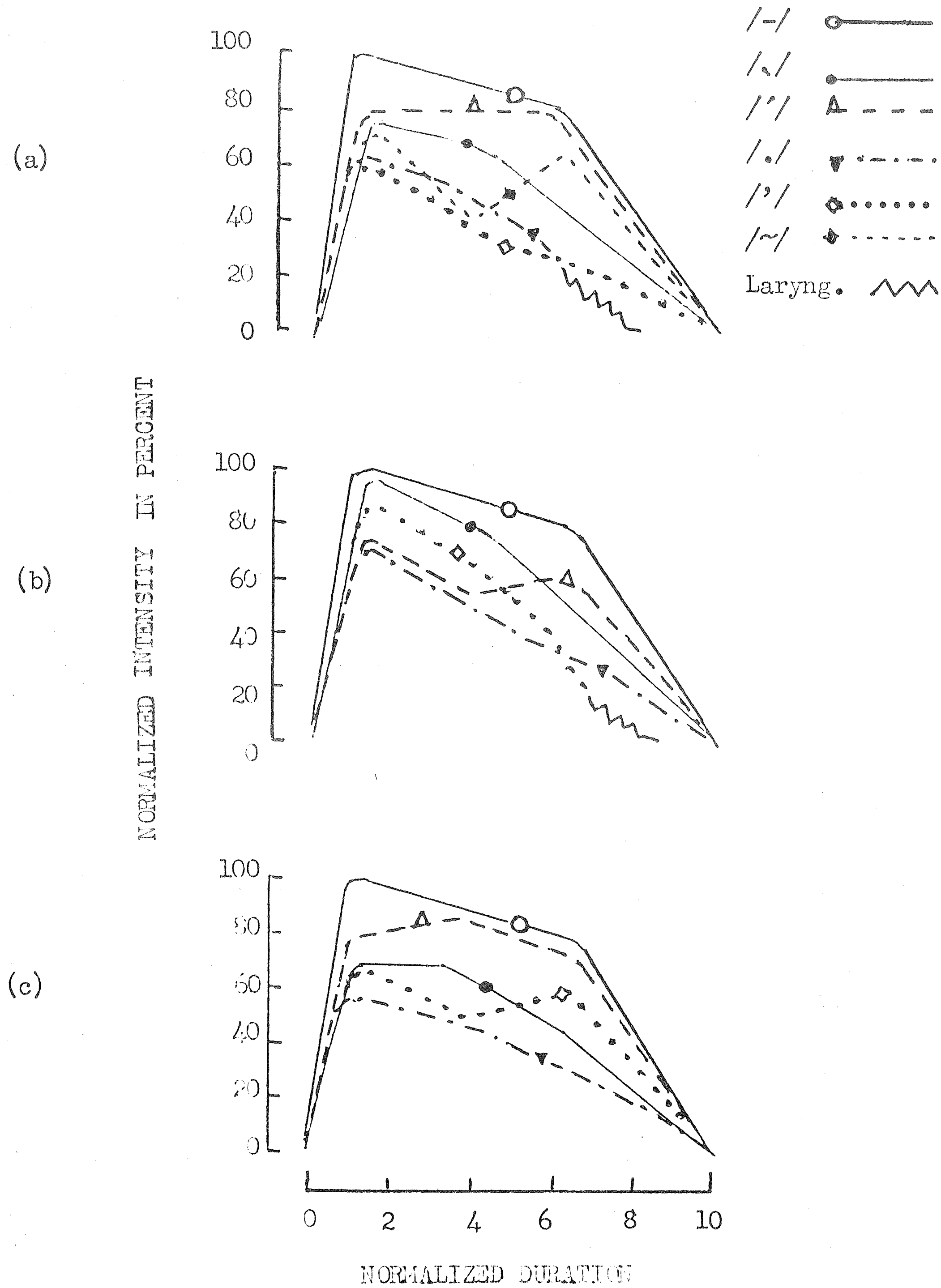


FIGURE 4.6 Intensity Curves (Synthetic Speech) of (a) NV, (b) CV and (c) SV Tones Used in Experiment IV (Normalized Intensity in Percent Plotted Against Normalized Duration)

Experiment III where the same intensity curve, that of the level tone, was used for all tones.

Figure 4.7 (following page) illustrates the overall patterns of recognition which look similar to those in Figure 4.4 (p 197) with the difference that recognition scores are generally lower. This suggests that the synthetic speech signals were not as good as they should be. Although some tones, e.g. the level tone and the rising tone in all dialects, gave more or less the same or even better recognition scores than in Experiment II, other tones did less well. The falling tone in particular suffered from a technical problem in tone synthesis which I could not solve at the time: the sudden cut-off of the signal at the end of each isolated syllable produced an impression similar to glottal closure which in this case sounded to me like the NV drop tone. Therefore recognition scores for the falling tone in both Experiments III and IV were less than 50% in most cases (cf Figure 4.5, p 198 and Figure 4.8, p 204 below.)

Some more facts can be learned from comparison between the results of Experiment III and Experiment IV concerning the role of intensity in tone perception. It would not be correct to infer from the general patterns illustrated in Figure 4.7 that intensity has different effects on tone perception in the different dialects: higher recognition scores for NV, and lower recognition scores for SV, in Experiment IV than in Experiment III, and no appreciable differences for CV.

A careful study of scores concerning individual tones,

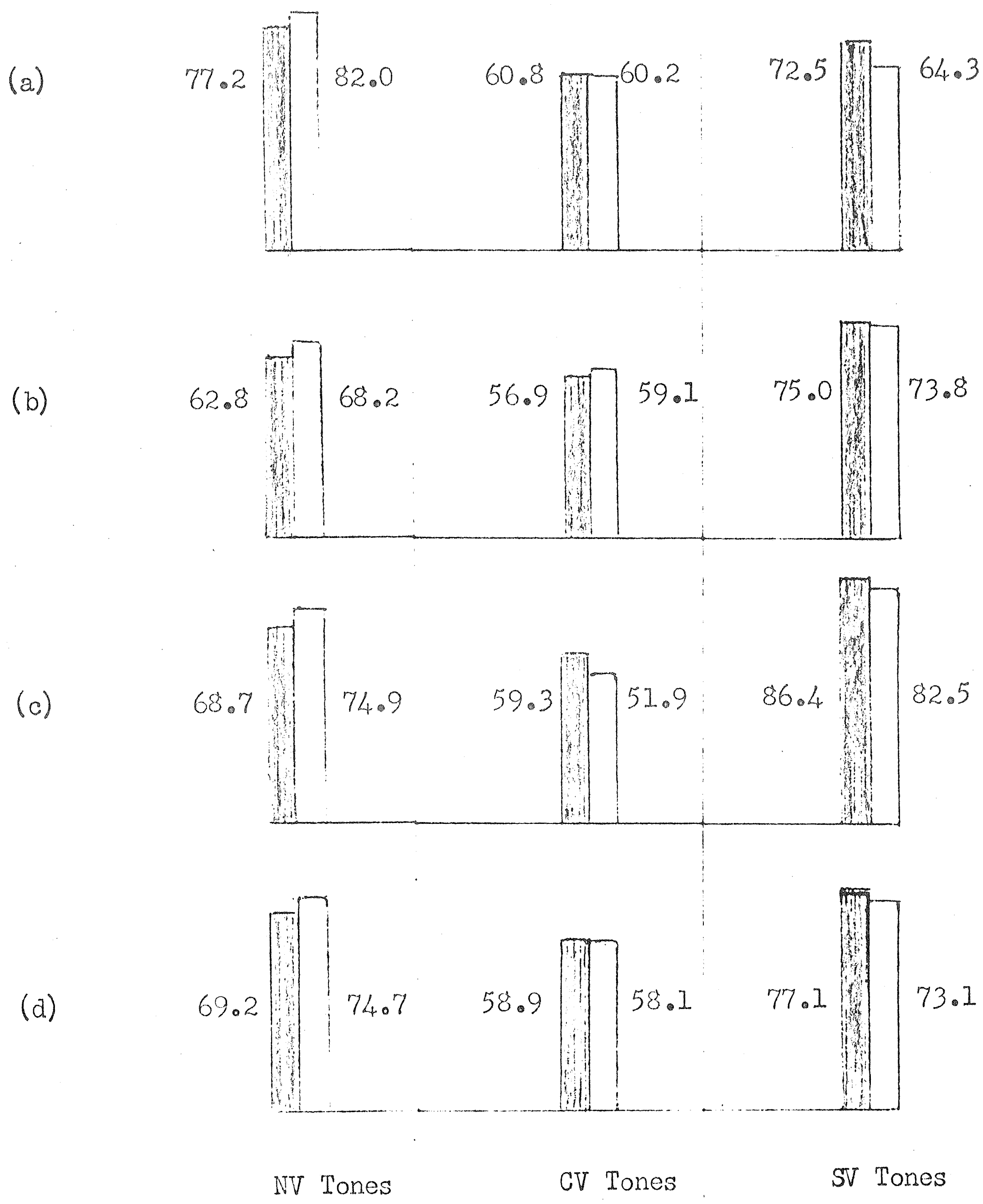


FIGURE 4.7 Overall Percentages of Correct Recognition of Synthetic Tones in Experiment III (Shaded Bars) and Experiment IV (Unshaded Bars) by (a) NV, (b) CV, (c) SV and (d) All V Subjects (Data From Tables 4.9 to 4.14)

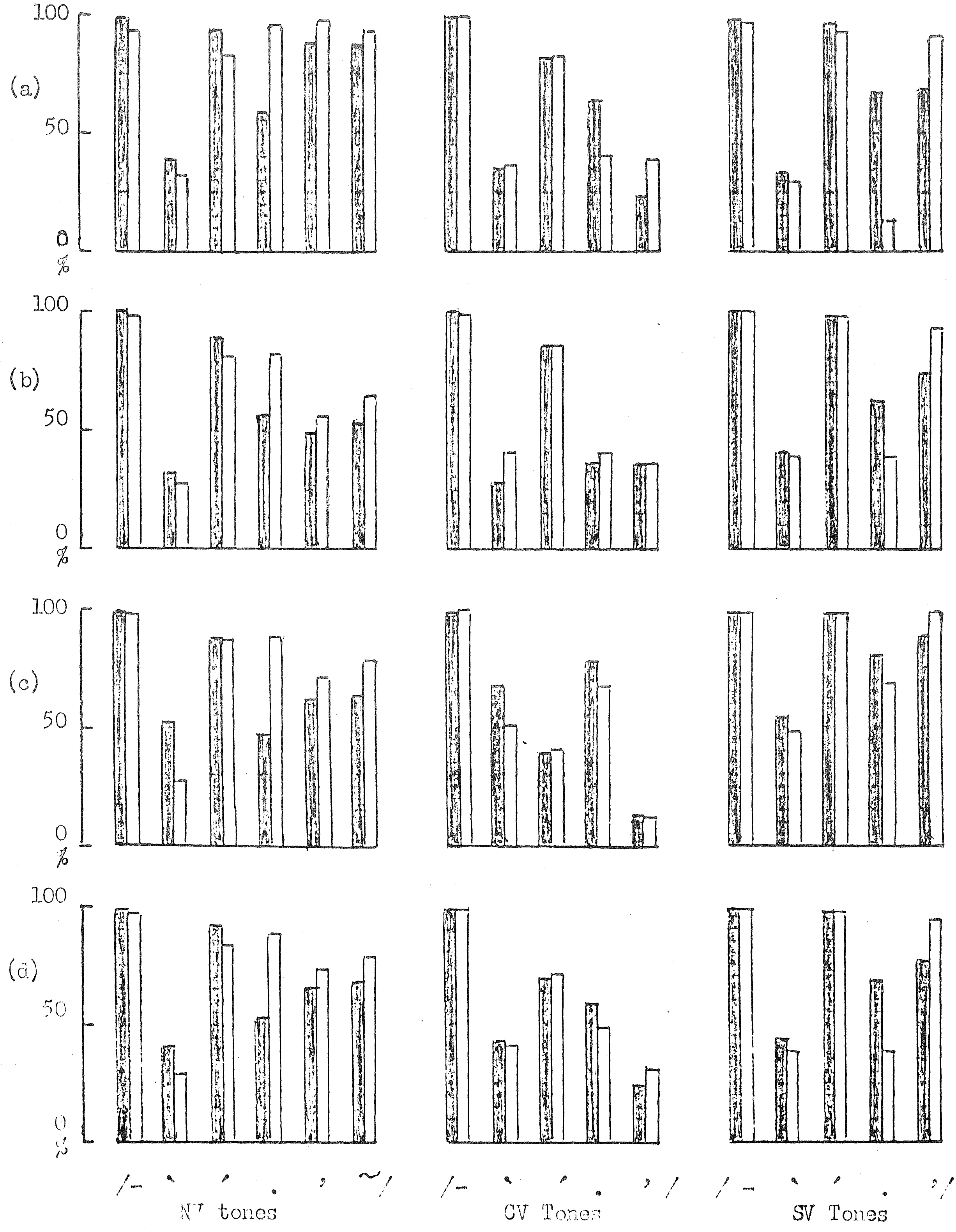


FIGURE 4.8 Percentages of Correct Recognition of Synthetic Tones in Experiment III (Shaded Bars) and Experiment IV (Unshaded Bars) by (a) NV, (b) CV, (c) SV and (d) All V Subjects (Data From Tables 4.9 to 4.14)

however, can offer a better understanding of the facts. Application of χ^2-tests on recognition scores of individual NV, CV and SV tones by all subjects, as graphically represented in part (d) of Figure 4.8, preceding page, indicates that differences are not the same for individual tones within each dialect. This is shown in Table 4.15 (next page) where actual recognition scores were given together with level of significance, and when there are significant differences, the direction of the differences, i.e. whether recognition is better (+) in Experiment IV, when Intensity variations were added, than in Experiment III, or worse (-).

Results of the tests show that (a) Intensity variations did not improve identification for the unidirectional tones (level, falling and rising) in all three dialects : either the differences were insignificant, or recognition got worse; (b) Intensity variations improved identification for the bidirectional tones (curve and broken) in all three dialects; and (c) they had variable effect on the identification of the drop tone : improvement for the falling creaky NV drop tone, but worsening for the CV and SV drop tones.

These facts suggest that for most cases intensity variations do not play a major role in tone perception. The exceptions to this are: the concave intensity contour enhances recognition for the NV broken tone and curve tone and the SV curve tone, possibly by reinforcing the existing concave Fo contour in its auditory effect, and better differentiating them from other tones; and the simulated creaky ending of the NV drop tone and the CV curve tone was better recognized when both intensity and Fo curves were represented by jumping, wavy lines in the input to the synthetizer.

TABLE 4.15 Differences in Recognition Scores of Synthetic Tones in Experiment III (Fo Variations Only) and Experiment IV (Fo and I Variations)*

Dia-lect	Parameters		Tones					
			Level	Falling	Rising	Drop	Curve	Broken
NV	Exp. III	Cor. R**	825	338	751	445	545	558
		Inc. R	9	491	81	388	292	282
	Exp. IV	Cor. R	750	225	644	688	571	605
		Inc. R	26	550	132	90	207	172
	Differ.	Signif.	0.01	0.01	0.01	0.01	0.01	0.01
		Direct.	-	-	-	+	+	+
CV	Exp. III	Cor. R	998	434	687	582	246	
		Inc. R	6	563	311	417	757	
	Exp. IV	Cor. R	930	377	667	448	283	
		Inc. R	4	548	267	485	644	
	Differ.	Signif.	ns	ns	ns	0.01	0.01	
		Direct.				-	+	
SV	Exp. III	Cor. R	984	432	967	687	765	
		Inc. R	10	569	30	297	231	
	Exp. IV	Cor. R	917	359	901	361	878	
		Inc. R	15	574	35	574	56	
	Differ.	Signif.	ns	0.05	ns	0.01	0.01	
		Direct.		-		-	+	

* Pooled scores for all NV, CV and SV Subjects.

** Cor. R : number of correct responses

Inc. R : number of incorrect responses

Signif.: level of significance according to χ^2-tests

Direct.: direction of differences; - means recognition scores are lower in Exp. IV than in Exp. III; + means recognition scores are higher in Exp. IV than in Exp. III.

Of particular interest is the fact that two NV synthetic tones, with Fo and intensity variations, yielded clearly improved recognition in Experiment IV. Scores were better than in Experiment III (with Fo variations only) and similar to those in Experiment II (real speech tones in isolated syllables.) This is shown below, with data extracted from Tables 4.6, 4.9 and 4.12.

	NV Drop Tone			NV Broken Tone		
Subjects :	NV	CV	SV	NV	CV	SV
Exp. II	94.6	80.5	93.8	92.0	54.7	58.4
Exp. III	58.8	55.4	46.5	86.5	53.0	64.0
Exp. IV	95.8	81.9	88.6	93.0	64.8	77.6

This suggests that the synthetic tones produced auditory impressions similar to the real tones. It is interesting to compare the spectrograms made from these synthetic tones used in Experiment IV and the same tones by Informant NM1 used in Experiment II (Figure 4.9, next page.) They show similar Fo contours; the simulated 'laryngealization' may not have the same acoustic quality as the real one, but the auditory impression could be close enough for listeners to identify the tones.

4.3.3 Tone Features and Perception

The problem of evaluating the relative importance of various tone features as perceptual cues is tackled here in a way different from the multidimensional scaling analysis described in Gandour 1978. My analysis proceeds from the conclusions reached

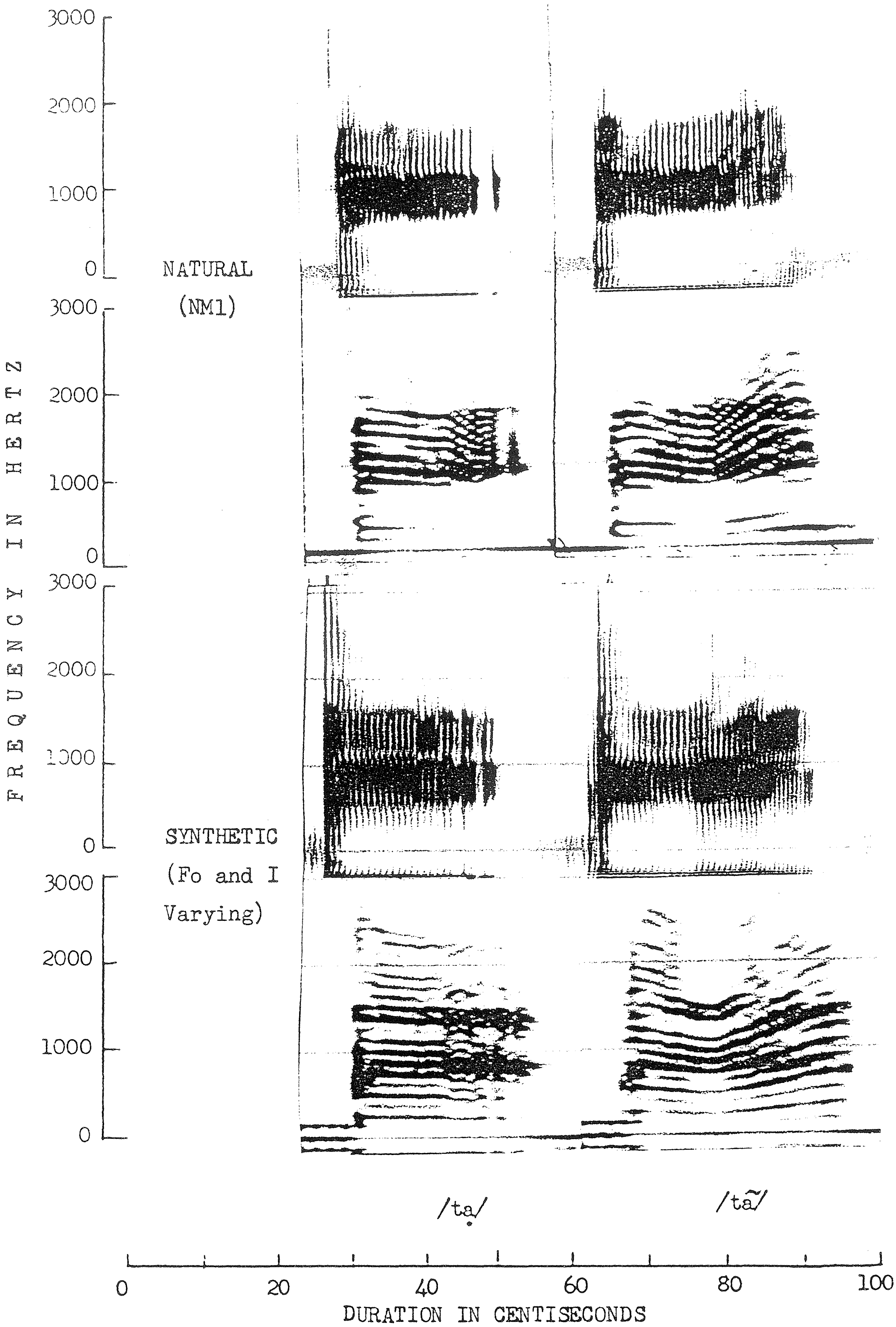


FIGURE 4.9 Wide-band and Narrow-band Spectrograms of NV Syllables with the Drop Tone and the Broken Tone (Natural and Synthetic)

so far and is based on the results of Experiment II, where no linguistic contextual cues were present as in Experiment I, and no unnatural factors were present as in Experiments III and IV. It revolves around the notion of misperception, defined as misidentification of a tone as another, due to misinterpretation of some tone features. The underlying assumption is that the confusion patterns reflected the native speakers' perception of tones in terms of similarity or dissimilarity in their features. Hence, it is presumed that in Experiment II conditions, when isolated tones were heard in a cross-dialect situation, tone perception was based on auditory cues interpreted as tone features and misperceptions usually occurred between two tones having some common features when some other features were misinterpreted as regards their nature or relative values. In this sense, a misperception is different from a simple error, when one tone is mistaken for another just by chance or for any reason in any situation, which is not of interest for our analysis.

For this purpose, all cases of misperception of 3.0% or more from the bottom boxes of Tables 4.6, 4.7 and 4.8 (overall response percentages by all Vietnamese subjects,) were picked out and presented in Table 4.16 (next page). The 3.0% cut-off was first chosen arbitrarily to take an appreciable but manageable number of cases for study, but later analysis justified that choice. Indeed, at 3.0% and above, no cases of misperception involving less than two common features were found, but the further inclusion of cases under 3.0% would have involved some tone pairs which share only one or no common features at all, presumably

TABLE 4.16 NV, CV and SV Tones Involved in Misperceptions in Experiment II *

No	(a) NV Tones	(b) CV Tones	(c) SV Tones
1	{ /ʼ/ ˧˩˧ -: 17.6 /./ ˨˩ˀ **	{ /ʼ/ ˧˩˧ˀ -: 11.6 /./ ˧˩	{ /ʼ/ ˨˩˦ -: 6.3 /./ ˨˩
2	{ 4.2 :-	{ 5.6 :-	{ 15.6 :-
3	{ /ʼ/ ˧˩˧ -: 12.7 /~/ ˧˥ˀ	/ʼ/ ˧˩˧ˀ -: 37.7 /~/ ˧˥ˀ	/ʼ/ ˨˩˦ -: 18.5 /~/ ˧˥ˀ
4	{ 28.6 :-	/ʼ/ ˧˩˧ˀ -: 5.6 /\/ ˧˩	/ʼ/ ˨˩˦ -: 15.0 /'/ ˧˥
5	/'/ ˧˥ -: 4.6 /ʼ/ ˧˩˧	/'/ ˧˥ -: 33.9 /ʼ/ ˧˩˧ˀ	/'/ ˧˥ -: 4.6 /-/ ˧
6	/'/ ˧˥ -: 6.2 /~/ ˧˥ˀ	/'/ ˧˥ -: 25.7 /~/ ˧˥ˀ	/./ ˨˩˦ -: 3.1 /'/ ˧˥
7	/./ ˨˩ˀ -: 4.5 /~/ ˧˥ˀ	/\/ ˧˩ -: 5.7 /-/ ˥	/./ ˨˩˦ -: 4.5 /~/ ˧˥ˀ
8	/\/ ˨˩ -: 4.3 /./ ˨˩ˀ	{ /\/ ˧˩ -: 6.8 /./ ˨˩	{ /\/ ˨˩ -: 15.2 /./ ˨˩
9	/ʼ/ ˧˩˧ -: 7.8 /\/ ˨˩	{ 48.8 :-	{ 5.4 :-
10		/-/ ˥ -: 3.0 /'/ ˧˥	/\/ ˨˩ -: 3.0 /ʼ/ ˨˩˦

* Data extracted from bottom boxes of Tables 4.6, 4.7 and 4.8 (see section 4.2) concerning misperceptions of 3.0 % or more by all NV, CV and SV subjects in Experiment II.

** { A -: 17.6 B
   4.2 :-

should be read "Tone A was perceived as B (17.6% of responses) and vice versa (4.2% of responses)".

constituting cases of simple error. In Table 4.16 and subsequent analysis, the sign -: (interpreted or perceived as) indicates direction of misperception, and the figures represent the percentages. The result is a total of 29 cases, 9 for NV and 10 each for CV and SV.

It is of interest to note that there are 12 cases (lines 1, 2, 3, and 8 in Table 4.16) which represent the same misperceptions (same tone pairs, same direction) in all three dialects, and eight other cases (lines 5, 6, 7 and 9) which involve the same misperceptions in two of the dialects. This is suggestive of the regularity of the patterns.

Furthermore, if one takes misperceptions of 10% or more as substantial, it may be noted that 13 out of those 29 cases represent substantial misperceptions (11.6% to 48.8%) and the most substantial ones occurred in CV tones. This agrees with the low recognition scores noted earlier in connection with CV tones.

A further analysis of tone features involved in those misperceptions is presented in Tables 4.17, 4.18 and 4.19 (following pages). The tone features involved were derived from categorization of physical phonetic parameters described in Ch. Two, and first introduced in Ch. One. Let us recall that they included average pitch (pitch for short) features [high] and [low], pitch contour (contour for short) features [falling] and [rising], and voice quality feature [creaky]. Other redundant features need not concern us here. From the above-mentioned three tables, in which features common to tone pairs involved in misperceptions are circled, the following observations can be made:

TABLE 4.17 Analysis of NV Tone Features Involved in Misperceptions in Experiment II*

No	Tone	Features					Perceived	Tone	Features				
		Hi	Lo	Fa	Ri	Cr	As		Hi	Lo	Fa	Ri	Cr
1	/ʼ/	(−)	(+)	(+)	+	−	−: *	/◦/	(−)	(+)	(+)	−	+ **
2							:−						
3	/ʼ/	−	+	(+)	(+)	−	−:	/~/	+	−	(+)	(+)	+
4							:−						
5	/´/	+	−	−	(+)	(−)	−:	/ʼ/	−	+	+	(+)	(−)
6	/´/	(+)	(−)	−	(+)	−	−:	/~/	(+)	(−)	+	(+)	+
7	/◦/	−	+	(+)	−	(+)	−:	/~/	+	−	(+)	+	(+)
8	/ˎ/	(−)	(+)	(+)	(−)	−	−:	/◦/	(−)	(+)	(+)	(−)	+
9	/ʼ/	(−)	(+)	(+)	+	(−)	−:	/ˎ/	(−)	(+)	(+)	−	(−)

* −: and ;− indicate directions of misperception.

** Features common to the two tones involved in misperceptions are circled. When the same tones are involved in both directions, tone marks and features are not repeated.

TABLE 4.18 Analysis of CV Tone Features Involved in Misperceptions in Experiment II*

No	Tone	Features					Perceived	Tone	Features				
		Hi	Lo	Fa	Ri	Cr	As		Hi	Lo	Fa	Ri	Cr
1	/ʼ/	⊖	-	⊕	+	+	-: *	/˳/	⊖	+	⊕	-	- **
2							:-						
3	/ʼ/	-	⊖	⊕	⊕	⊕	-:	/~/	+	⊖	⊕	⊕	⊕ * *
4	/ʼ/	⊖	⊖	⊕	+	+	-:	/ˎ/	⊖	⊖	⊕	-	-
5	/ʻ/	⊖	⊖	-	⊕	-	-:	/ʼ/	⊖	⊖	+	⊕	+
6	/ʻ/	-	⊖	-	⊕	-	-:	/~/	+	⊖	+	⊕	+
7	/ˎ/	-	⊖	+	⊖	⊖	-:	/-/	+	⊖	-	⊖	⊖
8	/ˎ/	⊖	-	⊕	⊖	⊖	-:	/˳/	⊖	+	⊕	⊖	⊖
9							:-						
10	/-/	+	⊖	⊖	-	⊖	-:	/ʻ/	-	⊖	⊖	+	⊖

* -: and :- indicate directions of misperception.

** Features common to the two tones involved in misperceptions are circled. When the same tones are involved in both directions, tone marks and features are not repeated.

* * Features of the broken tone /~/ are those of the NV dialect.

TABLE 4.19 Analysis of SV Tone Features Involved in Misperceptions in Experiment II*

No	Tone	Features					Perceived	Tone	Features				
		Hi	Lo	Fa	Ri	Cr	As		Hi	Lo	Fa	Ri	Cr***
1	{/ʾ/	⊖	-	⊕	⊕	⊖	-:*	/./	⊖	+	⊕	⊕	⊖**
2	{						:-						
3	/ʾ/	-	⊖	⊕	⊕	-	-:	/~/	+	⊖	⊕	⊕	+*
4	/ʾ/	-	⊖	+	⊕	⊖	-:	/ʹ/	+	⊖	-	⊕	⊖
5	/ʹ/	+	⊖	⊖	+	⊖	-:	/-/	-	⊖	⊖	-	⊖
6	/./	-	+	+	⊕	⊖	-:	/ʹ/	+	-	-	⊕	⊖
7	/./	-	+	⊕	⊕	-	-:	/~/	+	-	⊕	⊕	+
8	{/ˋ/	⊖	⊕	⊕	-	⊖	-:	/./	⊖	⊕	⊕	+	⊖
9	{						:-						
10	/ˋ/	⊖	+	⊕	-	⊖	-:	/ʾ/	⊖	-	⊕	+	⊖

* -: and :- indicate directions of misperception.

** Features common to the two tones involved in misperceptions are circled. When the same tones are involved in both directions, tone marks and features are not repeated.

* Features of the broken tone /~/ are those of the NV dialect.

*** The feature [creaky] does not normally apply in standard SV, but in cross-dialect perception, all SV tones should be marked [- creaky].

(a) All these misperceptions include at least two features common to the tone pairs involved; in fact out of the 29 cases, there are

9 cases involving 2 common features

11 cases involving 3 common features

9 cases involving 4 common features.

This supports the notion of misperception defined above and one may generalize that misperceptions often occur only when two or more common features are involved.

(b) In every one of these cases, at least one contour feature is common to both tones. In fact, there is appreciable difference in the number of cases involving different feature types:

	Common features in		
	Contour	Pitch	Voice Quality
No common features	0 case	6 cases	13 cases
One common feature	19 cases	12 cases	16 cases
Two common features	10 cases	10 cases	

The above figures mean that there is no case of systematic misperception occurring between two tones having no common contour features (falling, rising), while misperceptions occur more commonly between tones having one or two common contour features: 19 and 10 cases respectively. On the other hand, misperceptions often occur between tones having no common features in average pitch (high, low) or in voice quality (creaky): 6 and 13 cases respectively.

This fact suggests that the order of importance in Vietnamese tone perception is contour, pitch and voice quality. It also implies that contour features are the only obligatory common features involved in systematic misperceptions between two tones. In other words, misperceptions usually do not occur between tones of opposite contours (sharing no common features), namely rising versus falling, level versus concave, but often occur in all other cases, e.g. between level and rising (one common feature in contour: [- falling]), high and mid pitch, high and low pitch, creaky and non-creaky, etc.

An independent support for this analysis can be found in Miller 1961. The author found that the overall recognition score of NV tones in whispered speech was only 33% correct in a series of his tests (out of a total of 240 occurrences for the six tones.) He also made observations about a number of tones which were often interpreted as other tones and vice versa, and some tones which were rarely interpreted as other tones. A reanalysis of his observations (Miller 1961:14) along my lines is given in Table 4.20 (next page). It emerges from the table that only six of his cases coincide with mine (cf Table 4.17 and Table 4.20.) This is not surprising, given the differences between the two types of experiment. What is surprising is that all of 12 of Miller's cases described as occurring 'often' (15% or more) are fully conformable to the conditions given under (a) and (b) above, namely that they all include at least one common contour feature and at least another common feature of any type. On the other hand, the misperceptions described as 'rare' in his tests(mostly 2.5%) lend further support to my analysis: the tones involved have either only one single

TABLE 4.20 Analysis of NV Tone Features Involved in Misperceptions in Whispered Speech (Data from Miller 1961)

No	Tone	Features: Hi	Lo	Fa	Ri	Cr	Perceived As	Tone	Features: Hi	Lo	Fa	Ri	Cr
							(Often)						
<u>1</u>***	/ˎ/	(−)	(+)	(+)	(−)	−	−:*	/˳/	(−)	(+)	(+)	(−)	+ **
2	/-/	(−)	−	−	(−)	−	−:	/˳/	(−)	+	+	(−)	+
3	/~/	+	−	(+)	+	(+)	−:	/˳/	−	+	(+)	−	(+)
<u>4</u>	/ʾ/	(−)	(+)	(+)	+	−	−:	/˳/	(−)	(+)	(+)	−	+
<u>5</u>	/ʺ/	+	−	−	(+)	(−)	−:	/ʾ/	−	+	+	(+)	(−)
6	/-/	(−)	−	−	(−)	(−)	−:	/ˎ/	(−)	+	+	(−)	(−)
<u>7</u>	/~/	+	−	(+)	(+)	+	−:	/ʾ/	−	+	(+)	(+)	−
<u>8</u>							:−						
<u>9</u>	/ʾ/	(−)	(+)	(+)	+	(−)	−:	/ˎ/	(−)	(+)	(+)	−	(−)
10							:−						
11	/-/	−	(−)	(−)	−	(−)	−:	/ʺ/	+	(−)	(−)	+	(−)
12							:−						
							(Rarely)						
13	/-/	(−)	−	−	−	(−)	−:	/ʾ/	(−)	+	+	+	(−)
14	/ˎ/	−	+	(+)	−	−	−:	/~/	+	−	(+)	+	+
15							:−						
16	/ʺ/	+	−	−	+	−	−:	/˳/	−	+	+	−	+
17							:−						

* −: and :− indicate directions of misperception.

** Features common to the two tones involved in misperceptions are circled. When the same tones are involved in both directions, tone marks and features are not repeated.

*** Underlined numbers indicate same cases of misperceptions in Miller's experiments and mine (cf Table 4.17, p. 212).

common feature, or two common features not including contour ones, or no common features at all.

These facts seem to indicate that although pitch is not conveyed through whispered speech, the contour of tones is possibly faintly audible through intensity variations, which might become more salient and better perceived in whisper than in normal voice. And though this does not ensure correct recognition of the tones, it does favour misidentification between tones having some common contour features and prevent misidentification between those of markedly opposite contour features.

Another interesting fact concerns Miller's observation that "even the 80% correct identification of the low dipped [drop] tone... is not very convincing when we consider that listeners thought they heard the low dipped tone on 48 different occasions when it was not low dipped. Then, this tone, though appearing to be the easiest to hear, actually proved to be the most difficult as other tones were more readily confused with it." (Miller 1961:14). My analysis could give a better account for that. The NV drop tone is readily identified by its creaky ending. Even in whispered speech when relative pitch is not audible, this feature [+ creaky] could still be conveyed through other means such as intensity variations or a glottal stop, which could become a salient and sufficient feature in this situation. But even so it is not the only feature that characterizes this tone, and other tones which share some other feature with it, such as [+ falling], may still be readily confused with it, hence the fact described above.

The same analysis can account for the confusion between the rising tone and the falling tone in a local dialect in Nghi Loc district, Nghe Tinh province, reported in Bui Van Nguyen 1977. Speakers of this CV subdialect use the phrases /ká kɔ́ dũəj/ 'ka with a tail' and /kà kɔ́ kúəŋ/ 'ka with a stem' to distinguish between /ká/ 'fish' and /kà/ 'eggplant' when the context is of no help. Such confusion would not normally be possible in NV and SV where the two tones are maximally differentiated by opposition in all features except creaky voice

$$\text{NV \& SV /ˊ/} \begin{bmatrix} +\text{ high} \\ -\text{ low} \\ -\text{ falling} \\ +\text{ rising} \\ -\text{ creaky} \end{bmatrix} \qquad \text{NV \& SV /ˋ/} \begin{bmatrix} -\text{ high} \\ +\text{ low} \\ +\text{ falling} \\ -\text{ rising} \\ -\text{ creaky} \end{bmatrix}$$

In CV, the perceptual distance between these two tones is not so great with one variant of /ˊ/ (straight rising) and becomes minimal with the other variant (falling rising), both of wide use over the area:

$$\text{CV /ˊ/} \begin{bmatrix} -\text{ high} \\ -\text{ low} \\ -\text{ falling} \\ +\text{ rising} \\ -\text{ creaky} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} -\text{ high} \\ -\text{ low} \\ +\text{ falling} \\ +\text{ rising} \\ -\text{ creaky} \end{bmatrix} \qquad \text{CV /ˋ/} \begin{bmatrix} -\text{ high} \\ -\text{ low} \\ +\text{ falling} \\ -\text{ rising} \\ -\text{ creaky} \end{bmatrix}$$

It is not clear from the report whether the confusion occurs in some environment only or tonal merger has happened between these tones in this subdialect. But the analysis could predict both cases easily by a simple statement of feature change.

We may now summarize the observations made in this section by a more comprehensive statement which might be termed principles of tonal perception:

(1) Maximal opposition in tone features leads to greater perceptual distances and contrast between tones.

(2) Minimal opposition in tone features leads to lesser perceptual distances which may result in misperception and eventually to tonal merger.

(3) Various types of feature may play different roles in different languages and different situations. In Vietnamese, contour features play the most important role.

4.5 CONCLUSION

In this chapter, I have presented and discussed the results of perceptual experiments involving tones and subjects of the three major Vietnamese dialects.

The experiments show that tones in meaningful context can be readily identified (with high recognition scores) by speakers of the same dialect and other dialects. Tones in isolated syllables, predictably, have lower recognition scores and are better identified by speakers of the same dialect than by speakers of others. The special cases of CV tones, which have lower recognition scores than both NV and SV tones, even by CV speakers, may be due to a combination of various factors. These include their phonetic characteristics (less perceptual distances between tones regarding pitch levels and Fo range, more dissimilarity with NV and SV tones,) greater subdialectal variations within CV and possibly also their non-standard status. Recognition of synthetic tones show similar patterns, and the addition of intensity variations to Fo variations improves perception significantly only in a few cases, all involving concave contour or laryngealization, or both.

Analysis of the results shows that recognition patterns reflect the tone feature system proposed for Vietnamese. Correct perception and misperception are related to the similarity or dissimilarity in tone features. Contour features (falling, rising) are found to play a more important role than pitch features (high, low) and voice quality feature (creaky) in the identification of most tones, except for a few cases. This provides further support for the validity of the types of feature chosen for Vietnamese and their perceptual reality.

CHAPTER FIVE

A MODEL OF TONE PERCEPTION

5.0 *INTRODUCTION*

Proceeding from the results of perceptual experiments described in Ch. Four and the acoustic phonetic data presented in Ch. Two, I propose a theoretical model of the perceptual process for tone recognition in Vietnamese.

This model is based on the conclusions reached so far:

(a) Vietnamese tones in their diversity can be readily identified by speakers of different dialects, but the degree of correct recognition varies with the dialects, individual tones and listeners' background.

(b) Fo variation and laryngealization, perceived as pitch contour, relative pitch and voice quality, are the primary cues for tonal recognition. Intensity and duration may contribute secondarily to perception but seem to play no independent role in tone recognition.

(c) The proposed phonological tone features are found to be valid representation of psychological reality since the analysis of tone into such features provide insightful understanding of perceptual distances and misperceptions, tonal contrast and tonal confusion. This in turn helps to evaluate the relative importance of each type of feature, found to be of this order: contour, pitch and voice quality.

It is also based on the underlying assumption about the complex nature of Vietnamese tone as a physical and physiological reality, a perceptual entity and a phonological unit in the speech event. This complex nature has been shown in its various aspects and is reflected in different levels of representation related to each other by dynamic processes. The physical reality forms the basis for auditory perception, but this perception is also influenced by psychological and linguistic structures projected by the speaker-hearer's mind on what is perceived. As pointed out in Chomsky and Halle (1968:294), "a person's interpretation of a particular speech event is not determined merely by the physical properties constituting the event (...) what is perceived depends not only on the physical constitution of the signal but also on the hearer's knowledge of the language as well as on a host of extragrammatical factors." Thus there is dynamic interaction between various factors and not mechanical one-to-one correspondence. Therefore, the process of deriving phonological features from the primary acoustic data will involve different stages, which represent various degrees of abstraction and generalization.

The model has six components, taking as input the acoustic signals and giving as output the phonological tones. The components are: Physical Phonetic Parameters, Conversion Processes, Phonetic Features, Interpretation Rules, Phonological Features, and Influencing Factors, as represented schematically in Figure 5.1 (next

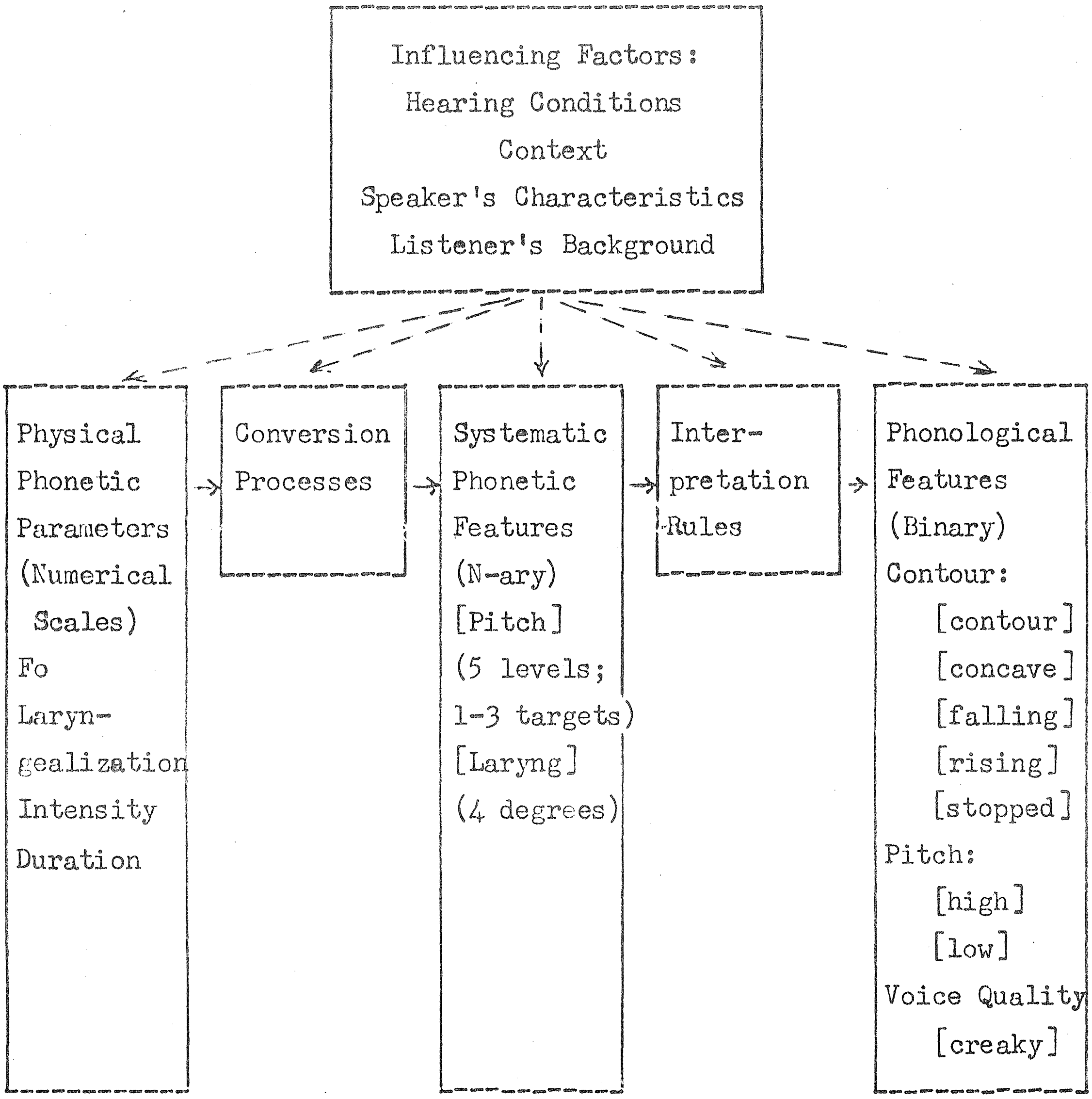


FIGURE 5.1 A Model of Tone Perception in Vietnamese

page), and explained in the subsequent sections.

5.1 *PHYSICAL PHONETIC PARAMETERS, PHONETIC FEATURES AND PHONOLOGICAL FEATURES*

The description of Vietnamese tones in Ch. Two and Ch. Three has provided for their analysis in terms of physical phonetic parameters of Fo, Laryngealization L, Intensity I and Duration D. These, in their normalized values obtained through normalization procedures described in 2.4.1, already represent a degree of abstraction when the fine acoustic details merge into broader categories.

At the systematic phonetic level, these parameters are converted into n-ary phonetic features, auditory and articulatory in character, which represent a further degree of abstraction by transforming the physical continuum into discrete units. The phonetic features of tone include pitch targets and laryngealization, which will be often referred to as [Pitch] and [Laryng] features for short. Intensity and duration are no longer represented, since they are predictable and their roles are not essential for tone perception, although information about them may be needed for better tone recognition in certain cases, as demonstrated in the experiments. A tone's pitch is represented by one, two or three figures, corresponding to one, two or three pitch targets described in Ch. Three, and representing onset, midpoint and endpoint pitch as the case arises. The degrees of laryngealization are indicated by figures or diacritics, and stopped tones are signalled by a small s appended to the pitch figures. The full notation will have the general form

$$\begin{bmatrix} \left\{ \begin{matrix} M \\ O(M)E(s) \end{matrix} \right\} & \text{Pitch} \\ L(L)(t) & \text{Laryng} \end{bmatrix}$$

in which the capital letters stand for values of the n-ary features (five levels for pitch targets and four degrees of laryngealization,) those in parentheses are optional, and (t) will specify the timing of laryngealization by the letter M (mid syllable) or E (end of syllable). Abbreviated forms include the use of diacritics and symbols in a modified version of Chao's (1930). Here is an example: the NV broken tone will be represented in its two variants (creaky voice or glottal stop at middle) in the following manner

$$\begin{bmatrix} 325 & \text{Pitch} \\ 2(3)M & \text{Laryng} \end{bmatrix} \quad \begin{matrix} \text{or} & \underset{\sim}{325} & \text{or} & \underset{\sim}{˧˨˥} \\ \text{or} & \underset{ʔ}{325} & \text{or} & \underset{ʔ}{˧˨˥} \end{matrix}$$

The phonological features of tone that are directly related to phonetic reality include the five features I have used to characterize all VN tones in the previous chapters. While phonetic features incorporate details which a linguist would be able to hear to describe different phonetic realizations of the same phonological tone, phonological features include only distinctive characteristics that native speakers need to differentiate the tones in actual communication.

As introduced in Ch. One (1.5.4.3), the phonological features also include 'contour,' 'concave' and 'stopped.' They are not needed in practical representation of each tone because they can be predicted by rules. The features 'concave' and 'contour' are predicted by two tone redundancy conventions as follows.

(TRC1) $$\begin{bmatrix} +\text{ falling} \\ +\text{ rising} \end{bmatrix} \rightleftarrows [+\text{ concave}]$$

and

(TRC2) $$\left\{ \begin{matrix} [+\text{ falling}] \\ [+\text{ rising}] \\ [+\text{ concave}] \end{matrix} \right\} \rightleftarrows [+\text{ contour}]$$

The double arrows mean that the conventions are applicable in both directions.

The feature 'stopped' is predicted by a phonological tone rule:

(PTR1) $$[T] \rightarrow \left\{ \begin{matrix} [+\text{ stopped}] \;/\; \_\_ \begin{bmatrix} C \\ -\text{ cont} \\ -\text{ son} \\ -\text{ voice} \end{bmatrix} \$ \\ [-\text{ stopped}] \;/\; \text{elsewhere} \end{matrix} \right\}$$

The application of these conventions and rules ensures that all bidirectional Vietnamese tones are concave and not convex, that a contour tone implies at least one of the contour features 'falling,' 'rising' or 'concave,' and that only stopped tones occur on syllables ending in a voiceless stop consonant. Thus, for example, besides the features specified in Table 5.1 (next page but one), predictable features will be added as follows.

Level tones	Rising tones	Stopped tones	Concave tones
- contour	+ contour	+ contour	+ contour
- concave	- concave	- concave	+ concave
- stopped	- stopped	+ stopped	- stopped

The processes of converting physical phonetic parameters into phonetic features, and then translating these features into phonological ones, will be described in the next sections.The results of these processes are given in Table 5.1 (next page).

5.2 *CONVERSION PROCESSES*

The processes described below are based on the assumption, supported by the analysis of perceptual test results in Ch. Four, that the two main cues for the recognition of Vietnamese tones are Fo and laryngealization. These can be directly related to the phonetic featuresof [Pitch] and [Laryng] by simple mathematical formulae which model the abstract auditory process. It is also assumed that the five levels of pitch targets correspond to the division of Fo ranges, in terms of $FD(\bar{F})$ percent, into three levels above and two levels below the mean $\bar{F}$. This means that positive $FD(\bar{F})$ values above or equal to zero will have pitch values of 3 to 5, and negative $FD(\bar{F})$ values will have pitch values of 1 or 2. This is based on the fact, which can be observed on Figure 2.17, in Ch. Two (p 92), that in each of the three dialects, the ratio between the average ranges of positive $FD(\bar{F})$ and negative $FD(\bar{F})$ is roughly 3:2.

The three following formulae will be sufficient to convert any $FD(\bar{F})$ value into Pitch value for all the tone targets.

$$\text{Pitch} = \text{Itg}\left(\frac{FD(\bar{F})}{PU} + 3\right) \qquad (C1)$$

TABLE 5.1 Correspondence between Physical Phonetic Parameters, Phonetic Features and Phonological Features of Standard NV, CV and SV Tones

Dialect & Tone	Fo in FD(F̄) %*			Phonetic Features*			Phonological Features					
	O	M	E	Pitch	Laryng	Abbrev.	Hi	Lo	Fa	Ri	Cr	
NV /-/	9	8	4	33	0	3 ˧	−	−	−	−	−	
/\`/	-7	-15	-19	21	0	21 ˨˩	−	+	+	−	−	
/´/	2	2	33	35	0	35 ˧˥	+	−	−	+	−	
/´s/	14	20	38	45s	0	45s ˦˥ s	+	−	−	+	−	**
/./	-2	-14	-15	21	2	21̰ ˨˩̰	−	+	+	−	+	
/.s/	-5	-13	-17	21s	0	21s ˨˩ s	−	+	+	−	−	
/ʼ/	-8	-26	-13	212	0	212 ˨˩˨	−	+	+	+	−	
/~/	4	-12	26	325	2	32̰5 ˧˨̰˥	+	−	+	+	+	
CV /-/	12	12	13	55	0	5 ˥	+	−	−	−	−	
/\`/	6	-1	-4	42	0	42 ˦˨	−	−	+	−	−	
/´/	-3	-8	10	24	0	24 ˨˦	−	−	−	+	−	
/´s/	0	-3	9	34s	0	34s ˧˦ s	−	−	−	+	−	
/./	0	-9	-9	31	0	31 ˧˩	−	+	+	−	−	
/.s/	3	-5	-8	31s	0	31s ˧˩ s	−	+	+	−	−	
/ʼ/	1	-9	-4	312	2	31̰2 ˧˩̰˨	−	−	+	+	+	
SV /-/	4	4	1	33	0	3 ˧	−	−	−	−	−	
/\`/	-9	-17	-18	21	0	21 ˨˩	−	+	+	−	−	
/´/	6	9	39	35	0	35 ˧˥	+	−	−	+	−	
/´s/	9	13	39	35s	0	35s ˧˥ s	+	−	−	+	−	
/./	-9	-19	-9	212	0	212 ˨˩˨	−	+	+	+	−	
/.s/	-7	-15	-12	21s	0	21s ˨˩ s	−	+	+	−	−	
/ʼ/	-5	-18	22	214	0	214 ˨˩˦	−	−	+	+	−	

* Fo parameters extracted from Table 2.15 (p 90). The M values represent either P3 or P4, whichever has lower Fo. The Laryng. values are the same for physical phonetic parameters and phonetic features.

** Stopped variants may have the same phonological features as basic forms or differ from them in one feature. They will be marked off by the feature 'Stopped'.

$$PU = \mathrm{Itg}\left(\frac{FD(\bar{F})max + 1}{3}\right) \qquad \text{if } FD(\bar{F}) \geqslant \emptyset \qquad (C2)$$

$$PU = \mathrm{Itg}\left(\frac{FD(\bar{F})min - 1}{-2}\right) \qquad \text{if } FD(\bar{F}) < \emptyset \qquad (C3)$$

where all values are expressed in integer digits (Itg),

PU is a pitch unit value determined for any $FD(\bar{F})$ range from which conversion applies,

$FD(\bar{F})max$ and $FD(\bar{F})min$ are the highest and lowest average $FD(\bar{F})$ respectively for each range, and + 1 and - 1 are correcting factors.

The pitch unit PU is the range of $FD(\bar{F})$ for each of the five pitch target levels in a tone system within the maximal $FD(\bar{F})$ and the minimal $FD(\bar{F})$ determined in a sample. The correcting factors +1 and - 1 in formulae (C2) and (C3) are needed to ensure that the five PUs add up to a value equal or greater than the whole range from which they are calculated. The correcting factor + 3 in formula (C1) ensures that the Pitch feature value always varies between 1 and 5 irrespective of the positive or negative values of the $FD(\bar{F})$.

Following are some examples of actual calculations.

The maximal and minimal $FD(\bar{F})$ of the SV dialect group are respectively 39 and -19, as can be gathered from Table 5.1 (preceding page). The PU for positive and negative $FD(\bar{F})$ calculated according to conversion formulae (C2) and (C3) are

$$PU = ltg\left(\frac{39 + 1}{3}\right) = 13 \text{ for positive } FD(\bar{F})$$

$$PU = ltg\left(\frac{-19 - 1}{-2}\right) = 10 \text{ for negative } FD(\bar{F})$$

The SV curve tone has the following $FD(\bar{F})$ at onset, midpoint and endpoint: -5, -18, 22 (from Table 5.1). Application of conversion formula (C1) gives the following results

$$O\ Pitch = ltg\left(\frac{-5}{10} + 3\right) = ltg(-0.5 + 3) = ltg(2.5) = 2$$

$$M\ Pitch = ltg\left(\frac{-18}{10} + 3\right) = ltg(-1.8 + 3) = ltg(1.2) = 1$$

$$E\ Pitch = ltg\left(\frac{22}{13} + 3\right) = ltg(1.6 + 3) = ltg(4.6) = 4$$

The Pitch values for the SV curve tone are thus determined to be [214] in systematic phonetic representation.

The application of formulae (C2) and (C3) to the $FD(\bar{F})$ values of the tones of all three dialects from Table 2.15 in Ch. Two gives the following data

	NV	CV	SV
$FD(\bar{F})max$	38	14	39
$FD(\bar{F})min$	-26	-9	-19
PU for positive $FD(\bar{F})$	13	5	13
PU for negative $FD(\bar{F})$	13	5	10

The application of formula (C1) to the onset, (midpoint) and endpoint $FD(\bar{F})$ of Vietnamese tones gives values for the pitch features in Table 5.1. Note that for the level tone, two values, for onset and endpoint, are used here, instead of one for midpoint target as proposed earlier in the previous section (5.2). The point is that the one-target analysis for the level tone remains valid for the tone production process and general description, but for the perceptual process, I have found that a two-value pitch representation for the level tone would be both more real and simpler to handle formally. Indeed, although the speaker needs only aim at one target when intending to produce a level tone, the hearer needs to compare at least two points to decide whether a tone has a level or a non-level contour. For the same reason, the input to conversion processes and interpretation rules should also include two values, at least, for any tone so that the same procedures could be applied without further complications.

Another point which needs elaboration is the procedure of determining a concave tone by scanning the midpoint values to see if there is one value which is lower by one pitch unit or more, as compared with onset and endpoint. The formula is

$$M = \emptyset \text{ if } O - \text{Mmin} < \text{PUmin} \ \underline{\text{or}} \ E - \text{Mmin} < \text{PUmin}$$
$$M \neq \emptyset \text{ if } O - \text{Mmin} \geq \text{PUmin} \ \underline{\text{and}} \ E - \text{Mmin} \geq \text{PUmin} \qquad \text{(C4)}$$

where M is the midpoint value to be determined and included in the systematic phonetic representation,

O and E are onset and endpoint values in $FD(\bar{F})$,

Mmin is the lowest midpoint value in $FD(\bar{F})$,

PUmin is the smaller of the two PU values for positive and negative $FD(\bar{F})$ in the tone system.

To take the same example given earlier in this section, the onset-midpoint and endpoint-midpoint differences in FD($\bar{F}$) of the SV curve tone are respectively

$(-5) - (-18) = 13$, and

$22 - (-18) = 40$.

Both these values are greater than 10, the PU for negative FD($\bar{F}$) for the SV group (see preceding page), therefore the midpoint Pitch value for this tone should be included, implying that it is a concave tone.

All the phonetic feature values given in Table 5.1 were obtained by application of Formulae (C1), (C2), (C3) and (C4). The only case where I decided to make an adjustment concerned the CV rising tone. Strict application of the formulae to the FD($\bar{F}$) values for this tone from Table 2.15 (Ch. Two) would yield a concave contour and an endpoint value of 5. But we know (see 2.4.3.2) that the concave variant of this CV tone may not be the basic form, and its endpoint could not be typically higher than its stopped variant. Therefore I decided that [24] would be more acceptable as the standard form. The mean Fo values for this tone had been biased by a few informants and in any case the values obtained were on the borderline between two pitch levels.

5.3 *TONE INTERPRETATION RULES*

These rules derive phonological tones from the phonetic tone features. They represent a further process of abstraction in which phonological features are mapped on to the phonetic shapes of tones.

I suggest that they are of the general form

A -: B if C

that is, A is interpreted as B if condition C is present. We need only four comprehensive rules to account for the interpretation of all tone features.

Contour Interpretation

$$(\text{TI1}) \quad [\text{O(M)E(s) Pitch}] \ -: \ \left\{ \begin{array}{ll} \begin{bmatrix} -\text{ falling} \\ -\text{ rising} \end{bmatrix} & \text{if } M = 0 \text{ and } E - 0 = \emptyset \\ \begin{bmatrix} +\text{ falling} \\ -\text{ rising} \end{bmatrix} & \text{if } M = \emptyset \text{ and } E - 0 \leqslant -1 \\ \begin{bmatrix} -\text{ falling} \\ +\text{ rising} \end{bmatrix} & \text{if } M = \emptyset \text{ and } E - 0 \geqslant 1 \\ \begin{bmatrix} +\text{ falling} \\ +\text{ rising} \end{bmatrix} & \text{if } M \neq \emptyset \end{array} \right\}$$

where M = Ø means no appreciable contour change at M, therefore no M value for the input,

M ≠ Ø means any M value, which signals a concave contour,

E - 0 represents the pitch differential between onset and endpoint, in 5-level values.

This rule captures the fact that Vietnamese listeners tend to perceive contour as a whole in terms of broad pitch movement, without regard to specific pitch values, as evidenced in the perceptual tests. Therefore, any difference of one or more levels would be interpreted as falling or rising contour. Redundancy conventions described in 5.1 will give the correct concave contour for the bidirectional tones.

Thus the level tones [33] and [55] will be interpreted as

[- falling, - rising] because the differences E - 0 are Ø in both cases. The NV and SV falling tones [21] and the CV falling tone [42] will be all interpreted as [+ falling, - rising] because the differences E - 0 are 1 - 2 = -1 and 2 - 4 = -2 respectively. All these tones are interpreted as non-concave because M = Ø, i.e. no contour changes at midpoint have been determined by formula (C4), applied during previous conversion processes.

Pitch Interpretation

(T12) [0(M)E(s) Pitch] -:

$$\left\{ \begin{array}{ll} \begin{bmatrix} + \text{high} \\ - \text{low} \end{bmatrix} & \text{if } \frac{0 + E}{2} \geqslant 4 \\ \begin{bmatrix} - \text{high} \\ - \text{low} \end{bmatrix} & \text{if } 2 < \frac{0 + E}{2} < 4 \\ \begin{bmatrix} - \text{high} \\ + \text{low} \end{bmatrix} & \text{if } \frac{0 + E}{2} \leqslant 2 \end{array} \right\}$$

where $\frac{0 + E}{2}$ is the evaluated value of average pitch.

This rule accounts for the fact that although five or even more levels of average pitch can be auditorily distinguishable, and may be phonemically distinctive in some languages, Vietnamese listeners would interpret them in terms of three levels of average pitch only. It also implies that the midpoint pitch, when representing a contour change, is more likely to contribute to the perception of contour rather than pitch, hence its exclusion from the evaluation formula for average pitch.

Thus, for example, the NV and SV rising tones will be interpreted as [+ high, - low] because their average pitch value is $\frac{3 + 5}{2} = 4$. The CV rising, falling and curve tones will be all interpreted as 'mid' [- high, - low], because their average pitch values are respectively $\frac{2 + 4}{2} = 3$, $\frac{4 + 2}{2} = 3$ and $\frac{3 + 2}{2} = 2.5$.

This rule also helps solve the problem of ambiguous average pitch level for the NV level and broken tones, which I discussed in Ch. Two (2.3.2b). For the level tone, application of conversion formula (C1) to the onset and endpoint FD($\bar{F}$), 9 and 4 respectively, gives the onset and endpoint Pitch values as $\text{Itg}(\frac{9}{13} + 3) = \text{Itg}(3.6) = 3$, and $\text{Itg}(\frac{4}{13} + 3) = \text{Itg}(3.3) = 3$, therefore the average pitch is also 3. This value places this tone firmly in the mid-pitch level.

For the broken tone [3̰25], exclusion of the midpoint value from the evaluation formula for average pitch also gives an unambiguous interpretation as [+ high, - low], because $\frac{3 + 5}{2} = 4$, which is consonant with auditory impression.

Voice Quality Interpretation

(T13) [L Laryng] -: $\begin{Bmatrix} \text{[- creaky]} & \text{if } L = 0 \text{ or } 1. \\ \text{[+ creaky]} & \text{if } L = 2 \text{ or } 3. \end{Bmatrix}$

where L = 0 means regular voicing

L = 1 means breathy voice

L = 2 means creaky voice

L = 3 means glottal closure.

This rule only formalizes the laryngeal facts described in Ch. Two. Clear voice [0 Laryng] and breathy voice [1 Laryng] will both be interpreted as [- creaky], since breathy voice occurs on some tones as a regular redundant feature with some speakers only. Creaky voice [2 Laryng] and glottal closure [3 Laryng] will both be interpreted as [+ creaky], since they occur alternatively on the laryngealized tones. In the standard dialects, only three tones, the NV drop tone, the NV broken tone and the CV curve tone will be interpreted as [+ creaky]. As noted before, in 2.3.6a, speakers of some local dialects in CV and SV also have some degree of creaky ending in their drop tones, therefore these may be marked [+ creaky] in their features. This does not alter their tone systems, because this feature is redundant in these cases and will only serve to enhance the perceptual distinguishability of these tones.

The application of the foregoing three tone interpretation rules will yield correct recognition of all NV, CV and SV tones in their respective systems, i.e. when the listener has access to the speaker's dialect. Otherwise, confusion may arise between different phonological tones having similar phonetic shapes in different dialects. For example, application of these rules to a tone having the phonetic shape [21] will yield

$$\begin{bmatrix} -\text{ high} \\ +\text{ low} \\ +\text{ falling} \\ -\text{ rising} \\ -\text{ creaky} \end{bmatrix}$$

which is the feature composition of the NV and SV falling tone and the CV drop tone. This explains why in Experiment II, when subjects had to identify tones in isolated syllables and to guess at the dialect they belonged to, the CV drop tone was identified as falling tone at the rate of 57.8% by NV subjects, 26.3% by CV subjects and 64.9% by SV subjects (see Table 4.7, Ch. Four).

There is one more point concerning these rules. As they are, they will yield exactly the feature compositions of all NV, CV and SV tones as specified in Table 5.1. They predict that the feature compositions of the stopped rising tones in all dialects and the CV stopped drop tone will be the same as those of the corresponding non-stopped tones. However, the feature compositions of the NV and SV stopped drop tones will differ from their non-stopped counterparts in one feature, namely in the values of 'creaky' and 'rising' respectively. They are the same as those of the NV and SV falling tones, and that is why non-Vietnamese listeners would interpret them as falling tones if they are not acquainted with the Vietnamese phonological system.

If one adopts a surface phonemic solution by recognizing the stopped tones as separate tonemes, the problem will be solved by applying rule PTR1 (introduced in 5.2 above) which adds the feature [+ stopped] to all these tones and makes them distinctive from the other tones. This results in an eight-tone system for NV and seven-tone systems for CV and SV. This solution was favoured by some earlier Vietnamese grammarians.

However, an underlying six-tone system is favoured in the official orthography and by most authors cited in 1.1, mainly on the grounds of phonetic similarity, complementary distribution and economy. I also favour this solution and would add the argument that the differences in phonetic details between the stopped and non-stopped tones regarding Fo, lack of laryngealized end or of rising end in some tones, and shorter duration, can be explained as conditioning by the presence of the voiceless final stop (see 3.1.2, Ch. Three). Therefore the feature 'stopped' is not an underlying one and can be predicted from this presence by a phonological rule.

In this case, another tone interpretation rule is needed to handle the NV and SV stopped drop tone, to ensure that they are interpreted as underlying drop tones and not underlying falling tones.

Stopped Tone Feature Reinterpretation

$$\text{(T14)} \quad [\text{O(M)Es Pitch}] \;\text{-:}\; \left\{ \begin{matrix} \begin{bmatrix} +\ \text{NV} \\ +\ \text{creaky} \end{bmatrix} \\ \begin{bmatrix} +\ \text{SV} \\ +\ \text{rising} \end{bmatrix} \end{matrix} \right\} \quad \left\{ \begin{matrix} \text{if} \quad \text{s present, and} \\ \text{E - O} \leq -1 \end{matrix} \right\}$$

This is a special rule, because unlike the others, it does not purport to represent the reality of tone perception alone, but that of tone-segment relationship at syllable level. Listeners are not supposed to perceive that the stopped variants have the same features as the non-stopped forms in these cases, but are supposed to derive this information from the presence of the voiceless stop at the end of the syllable.

Thus, the NV stopped drop tone [21s], having E - O = 1 - 2 = -1 (i.e. falling contour) will be reinterpreted as [+ creaky] and having the same underlying features as the non-stopped drop tone [21̰], i.e.

$$\begin{bmatrix} -\ \text{high} \\ +\ \text{low} \\ +\ \text{falling} \\ -\ \text{rising} \\ +\ \text{creaky} \end{bmatrix}$$

Similarly, the SV stopped drop tone [21s] will be reinterpreted as [+ rising] and having the same features as the SV non-stopped drop tone [212], i.e.

$$\begin{bmatrix} -\ \text{high} \\ +\ \text{low} \\ +\ \text{falling} \\ +\ \text{rising} \\ -\ \text{creaky} \end{bmatrix}$$

This rule should be ordered after TI1 and TI3 because if these rules apply after it, they will nullify its effect and give back the phonetically-based features.

This aspect of feature reinterpretation was not tested in my perceptual experiments because the results would be predictable. Natural stopped tones inserted in between non-stopped tones would be singled out easily by native speakers. I had not thought of a way of removing the final consonant while keeping the original Fo curve of a stopped tone by some technical device. If this is possible, I would predict that the results would be : all stopped rising tones and the CV stopped drop tone would be recognized by speakers of respective dialects as non-stopped tones of the same names, but the NV and SV stopped drop tones would be identified as falling tones. This is also what my first three tone interpretation rules (TI1, TI2 and TI3) would predict. It would be interesting to have a special experiment designed to test these predictions.

The assumption that all this has reality in the Vietnamese phonological system not only comes from my intuition as a native speaker, but is based on the following two facts: (a) As the six-tone system is explicitly recognized in the official orthography, all literate Vietnamese have used it without difficulty. No reports have been made about Vietnamese school-children making mistakes in using tone marks for the stopped tones. The usual mistakes concern the use of the diacritics [’] and [~] for the curve and broken tones, because they have historically merged in CV and SV. (b) Rhyming and alternation patterns in Vietnamese poetry and folk-songs exhibit a clear dichotomy between the traditionally called 'even rhymes', where only the level and falling tones can occur, and 'uneven rhymes', where all other tones, including the stopped tones, can occur. This suggests that in the native speakers' minds, the stopped tones belong to the same class as the rising tones and drop tones and share underlying features with these and not with the falling tone.

5.4 EXAMPLES OF COMPLETE DERIVATION

The sequential application of normalization procedures, conversion formulae and tone interpretation rules will yield correct feature compositions for all NV, CV and SV tones as given for the standard forms in Table 5.1. Note that with the exception of T14 where the diacritic features NV and SV appear, all the processes involved apply in all dialects in a uniform way. This points to the underlying unity of the three dialects in tonal features.

Following are examples of how the phonological features can be derived, step by step, from the acoustic parameters, leading to the identification of specific tones.

(a) The NV Rising Tone

Fo at six timepoints : 198 194 197 217 246 257 Hz

Laryng : 0

Application of normalization formula (N2) gives the following values of $FD(\bar{F})$, in percent, with $\bar{F}$ = 193 Hz for the NV group:

2 0 2 12 27 33 %

An example of how the endpoint $FD(\bar{F})$ value 33 is obtained:

$$(N2)\quad FD(\bar{F}) = Itg(\frac{Fi - \bar{F}}{\bar{F}} \times 100) = Itg(\frac{257 - 193}{193} \times 100)$$

$$= Itg(33.16) = 33\ \%$$

The PU values are determined for NV by applying formulae (C2) and (C3), knowing that the highest and lowest $FD(\bar{F})$ values for the NV informant group under study are 38 and -26 respectively.

$$(C2)\quad PU = Itg(\frac{38 + 1}{3}) = 13 \text{ for positive } FD(\bar{F}).$$

$$(C3)\quad PU = Itg(\frac{-26 - 1}{-2}) = 13 \text{ for negative } FD(\bar{F}).$$

Application of formula (C4) determines that no M value is needed for this tone:

$$(C4)\quad 0 - Mmin = 2 - 2 = \emptyset < PUmin\ (13).$$

Application of conversion formula (C1) to the onset and endpoint $FD(\bar{F})$ values will yield the Pitch values as follows:

$$(C1)\quad 0\ Pitch = Itg(\frac{FD(\bar{F})}{PU} + 3) = Itg(\frac{2}{13} + 3) = 3.$$

$$E\ Pitch = Itg(\frac{33}{13} + 3) = 5.$$

Application of the contour interpretation rule to the phonetic feature [35 Pitch] will yield the contour features [- falling, + rising]:

(T11) [35 Pitch] -: $\begin{bmatrix} - \text{ falling} \\ + \text{ rising} \end{bmatrix}$ because $M = \emptyset$ and $E - 0 = 2$.

Application of the pitch interpretation rule will yield the pitch features [+ high, - low] :

(T12) [35 Pitch] -: $\begin{bmatrix} + \text{ high} \\ - \text{ low} \end{bmatrix}$ because $\frac{0 + E}{2} = \frac{3 + 5}{2} = 4$.

Application of the voice quality interpretation rule will yield the feature [- creaky] :

(T13) [0 Laryng] -: [- creaky]

Thus, the final result is $\begin{bmatrix} + \text{ high} \\ - \text{ low} \\ - \text{ falling} \\ + \text{ rising} \\ - \text{ creaky} \end{bmatrix}$

This is the feature composition of the NV rising tone, with other redundant features (contour, concave) predicted by tone redundancy conventions described in 5.2 supra.

(b) <u>The SV Stopped Drop Tone</u>

Its derivation will be given below in successive applications of formulae and rules, with brief explanations only when necessary. The SV group has a maximal $FD(\bar{F})$ of 39 and a minimal $FD(\bar{F})$ of -19.

Fo at four timepoints: 170 159 155 160 Hz

Laryng : 0

(N2) $FD(\bar{F})$, with $\bar{F}$ = 183 Hz: -7 -13 -15 -12 %

(C2) $PU = \text{Itg}\left(\frac{39 + 1}{3}\right) = 13$ for positive $FD(\bar{F})$.

(C3) $PU = \text{Itg}\left(\frac{-19 - 1}{-2}\right) = 10$ for negative $FD(\bar{F})$.

(C4) Test for midpoint value

$E - Mmin = (-12) - (-15) = 3 \leqslant PUmin\ (10).$

Therefore $M = \emptyset$, no concave contour.

(C1) $0\ Pitch = Itg(\frac{-7}{10} + 3) = 2$

$E\ Pitch = Itg(\frac{-12}{10} + 3) = 1$

(T11) [21s Pitch] -: $\begin{bmatrix} + \text{falling} \\ - \text{rising} \end{bmatrix}$ because $M = \emptyset$ and $1 - 2 = -1$.

(T12) [21s Pitch] -: $\begin{bmatrix} - \text{high} \\ + \text{low} \end{bmatrix}$ because $\frac{2 + 1}{2} = 1.5 < 2$.

(T13) [0 Laryng] -: [- creaky]

(T14) [21s Pitch] -: $\begin{bmatrix} + \text{SV} \\ + \text{rising} \end{bmatrix}$ because s present, and $1 - 2 = -1$.

Thus, this last rule changes the feature [- rising] yielded by rule T11 to [+ rising] and gives this stopped drop tone the same underlying feature composition as the SV non-stopped drop tone:

$$\begin{bmatrix} - \text{high} \\ + \text{low} \\ + \text{falling} \\ + \text{rising} \\ - \text{creaky} \end{bmatrix}$$

(c) <u>The CV Curve Tone by Informant CF4</u>.

This informant has a mean $\bar{F}$ of 190 Hz, a highest mean Fo of 232 Hz and a lowest mean Fo of 166 Hz, therefore $FD(\bar{F})max$ is 22% and $FD(\bar{F})min$ is -12%.

Fo at six timepoints 193 187 177 166 197 204 Hz

Laryng : 2

(N2) FD($\bar{F}$) 1 -1 -6 -12 3 7 %

(C2) PU = Itg$\left(\frac{22 + 1}{3}\right)$ = 7 for positive FD($\bar{F}$).

(C3) PU = Itg$\left(\frac{-12 - 1}{-2}\right)$ = 6 for negative FD($\bar{F}$).

(C4) Test for midpoint value:

O - Mmin = 1 - (-12) = 13 > PUmin(6), and

E - Mnin = 7 - (-12) = 19 > PUmin (6).

Therefore M $\neq \emptyset$, there is concave contour.

(C1) O Pitch = Itg$\left(\frac{1}{7} + 3\right)$ = Itg(3.14) = 3.

M Pitch = Itg$\left(\frac{-12}{6} + 3\right)$ = Itg(1.0) = 1.

E Pitch = Itg$\left(\frac{7}{7} + 3\right)$ = Itg(4.0) = 4.

Thus, the phonetic representation for this tone is [314̰]

(T11) [314 Pitch] -: $\begin{bmatrix} + \text{falling} \\ + \text{rising} \end{bmatrix}$ because M $\neq \emptyset$.

(T12) [314 Pitch] -: $\begin{bmatrix} - \text{high} \\ - \text{low} \end{bmatrix}$ because $\frac{3 + 4}{2}$ = 3.5 (<4,>2)

(T13) [2 Laryng] -: [+ creaky]

This example shows that although the curve tone by Informant CF4 has a slightly different phonetic shape than the standard CV curve tone given in Table 5.1, viz. [312̰], it will be identified as such because the phonological feature composition is the same:

$$\begin{bmatrix} - \text{high} \\ - \text{low} \\ + \text{falling} \\ + \text{rising} \\ + \text{creaky} \end{bmatrix}$$

5.5 *THE INFLUENCING FACTORS*

The correct identification of tones depends not only on the processes described above, but also on a number of other factors which may influence them. We know from the study of tone variation that the phonetic shape of a tone may deviate a good deal from the basic tone form. From the perceptual point of view, the fact that different variants can be identified as the same tone and one tone may be misinterpreted as another can be attributed to the following factors which interact with the conversion processes and interpretation rules to yield correct or incorrect recognition:

(a) Hearing conditions may influence perception at the very start of the process through a series of factors: the clearness of the acoustic signals, the presence of noise or other speech occurring at the same time, the physiological and psychological readiness of the listener to sustain attention, etc.

(b) Context, both linguistic and extra-linguistic, is a major factor which enhances perception in any situation. This is evidenced by the differences between the results of my experiments I and II, and those of Miller 1961 who found that whispered tones in meaningful utterances could be highly recognizable, whereas whispered isolated tones could hardly be.

(c) Speaker's characteristics can also enhance or impede recognition, depending on whether they follow or deviate much from the norms, and whether information is available on them. For instance, Abramson 1976 concluded from perceptual tests on Thai tones that "phonemic tones free of a linguistic context are better identified when the listener has access to the speaker's tone space."

(d) The listener's background can also play a part in the interpretation process. It reflects the listener's "relative perceptual saliency" (Gandour 1978), as exemplified by the differences in recognition scores between NV, CV and SV subjects in my experiments.

5.6 THE PERCEPTUAL MODEL AND TONE VARIATION

The examples of derivation given in 5.4 have shown that the model is able to derive correct phonological features from the acoustic material for standard tone forms. Variation in phonetic detail will normally be resolved through the derivation process. However, if the model is applied to some variant forms in local dialects or in actual speech that deviate considerably from the norms, some modifications or refinements will be needed.

For example, to accommodate the fact that the CV rising tone has two variants, one mid-rising and one mid-concave, as noted earlier, its phonological feature composition would have to include an alternative value $\pm$ (plus or minus) for the feature 'rising.' It would be distinguishable from the CV curve tone by either two features (falling, creaky) or only one (creaky), as their phonological representations will be

$$\text{CV /ˊ/} \begin{bmatrix} -\ \text{high} \\ -\ \text{low} \\ \pm\ \text{falling} \\ +\ \text{rising} \\ -\ \text{creaky} \end{bmatrix} \quad \text{and CV /’/} \begin{bmatrix} -\ \text{high} \\ -\ \text{low} \\ +\ \text{falling} \\ +\ \text{rising} \\ +\ \text{creaky} \end{bmatrix}$$

The problem will be more complex if the model is to derive underlying features from tone variants in actual speech, which may have undergone some tone sandhi process. It is beyond the scope of this study to investigate the matter, but the problem would not be so great because, according to some authors, tone sandhi does not occur in Vietnamese (see Han and Kim 1972 and Tran Huong Mai 1969, quoted in 3.1.4.) However,data in my possession on some connected utterances show that the variations are sometimes great enough to require adjustments in the derivation process if correct identification is to be ensured. The model certainly needs greater adjustments and additions if it is to handle the perception of tone in languages where tone sandhi and tone assimilation in various forms occur regularly.

APPENDIX TO CHAPTER FIVE

SUMMARY OF FORMULAE AND RULES

The formulae and rules needed to derive phonetic and phonological features from acoustic data can be found at various places in Ch. Two and Ch. Five. The following summary lists them in the order they should normally be applied. The page number refers to where they are first introduced and explained.

Normalization Formula

(N2) Fo Differential relative to the mean $\bar{F}$ (p 82)

$$FD(\bar{F}) = Itg(\frac{Fi - \bar{F}}{\bar{F}} \times 100)$$

Conversion Formulae

(C2) Pitch Unit for positive $FD(\bar{F})$ (p 230)

$$PU = Itg(\frac{FD(\bar{F})max + 1}{3}) \text{ if } FD(\bar{F}) \geq 0$$

(C3) Pitch Unit for negative $FD(\bar{F})$ (p 230)

$$PU = Itg(\frac{FD(\bar{F})min - 1}{-2}) \text{ if } FD(\bar{F}) < 0$$

(C4) Test for midpoint value (p 232)

$M = \emptyset$ if $O - Mmin < PUmin$ or $E - Mmin < PUmin$

$M \neq \emptyset$ if $O - Mmin \geq PUmin$ and $E - Mmin \geq PUmin$

(C1) Conversion of $FD(\bar{F})$ into Pitch value (p 228)

$$Pitch = Itg(\frac{FD(\bar{F})}{PU} + 3)$$

Tone Interpretation Rules

(TI1) Contour Interpretation (p 234)

$$[O(M)E(s)\ \ \text{Pitch}] \ -: \left\{ \begin{array}{lll} \begin{bmatrix} -\ \text{falling} \\ -\ \text{rising} \end{bmatrix} & \text{if} & \begin{array}{l} M = \emptyset \text{ and} \\ E - O = \emptyset \end{array} \\ \begin{bmatrix} +\ \text{falling} \\ -\ \text{rising} \end{bmatrix} & \text{if} & \begin{array}{l} M = \emptyset \text{ and} \\ E - O \leqslant -1 \end{array} \\ \begin{bmatrix} -\ \text{falling} \\ +\ \text{rising} \end{bmatrix} & \text{if} & \begin{array}{l} M = \emptyset \text{ and} \\ E - O \geqslant 1 \end{array} \\ \begin{bmatrix} +\ \text{falling} \\ +\ \text{rising} \end{bmatrix} & \text{if} & M \neq \emptyset \end{array} \right\}$$

(TI2) Pitch Interpretation (p 235)

$$[O(M)E(s)\ \ \text{Pitch}] \ -: \left\{ \begin{array}{lll} \begin{bmatrix} +\ \text{high} \\ -\ \text{low} \end{bmatrix} & \text{if} & \frac{O + E}{2} \geqslant 4 \\ \begin{bmatrix} -\ \text{high} \\ -\ \text{low} \end{bmatrix} & \text{if} & 2 < \frac{O + E}{2} < 4 \\ \begin{bmatrix} -\ \text{high} \\ +\ \text{low} \end{bmatrix} & \text{if} & \frac{O + E}{2} \leqslant 2 \end{array} \right\}$$

(TI3) Voice Quality Interpretation (p 236)

$$[L\ \ \text{Laryng}] \ -: \left\{ \begin{array}{lll} [-\ \text{creaky}] & \text{if} & L = 0 \text{ or } 1 \\ [+\ \text{creaky}] & \text{if} & L = 2 \text{ or } 3 \end{array} \right\}$$

(TI4) Stopped Tone Feature Reinterpretation (p 239)

$$[\mathrm{O(M)Es\ \ Pitch}] \; -: \; \left\{ \begin{matrix} \begin{bmatrix} +\ \mathrm{NV} \\ +\ \mathrm{creaky} \end{bmatrix} \\ \\ \begin{bmatrix} +\ \mathrm{SV} \\ +\ \mathrm{rising} \end{bmatrix} \end{matrix} \quad \text{if} \quad \begin{matrix} \text{s present and} \\ \mathrm{E} - \mathrm{O} \leqslant -1 \end{matrix} \right\}$$

<u>Other Tone Rules</u>

(TRC1) Redundancy Convention for Concavity (p 227)

$$\begin{bmatrix} +\ \mathrm{falling} \\ +\ \mathrm{rising} \end{bmatrix} \rightleftarrows [+\ \mathrm{concave}]$$

(TRC2) Redundancy Convention for Contourity (p 227)

$$\left\{ \begin{matrix} [+\ \mathrm{falling}] \\ [+\ \mathrm{rising}] \\ [+\ \mathrm{concave}] \end{matrix} \right\} \rightleftarrows [+\ \mathrm{contour}]$$

(PTR1) Stopped Feature Assignment (p 227)

$$[\mathrm{T}] \rightarrow \left\{ \begin{matrix} [+\ \mathrm{stopped}] \ / \ \underline{\quad} \begin{bmatrix} \mathrm{C} \\ -\ \mathrm{cont} \\ -\ \mathrm{son} \\ -\ \mathrm{voice} \end{bmatrix} \$ \\ \\ [-\ \mathrm{stopped}] \ / \ \text{elsewhere} \end{matrix} \right\}$$

CONCLUSION

This thesis has been an attempt to describe Vietnamese tones in the three major dialects at different levels of representation.

I have shown that at the physical phonetic level, the acoustic phonetic properties of tone can be described quantitatively and normalized into the physical phonetic parameters of Fo, Intensity, Duration and Laryngealization. These are useful for comparing and assessing similarities and differences between speakers and dialects. Their nature can be understood by referring to the laryngeal phonation process in which there is complex and dynamic interaction between extrinsic and intrinsic factors. The data established Vietnamese as a "contour tone language with register overlap". Four types of Fo contour and three average Fo levels were found necessary to characterize all Vietnamese tones. Laryngealization was also found to be distinctive or redundant features on some tones.

At the systematic phonetic level, the tones can be described in terms of phonetic features representing different pitch targets and varying degrees of laryngealization. Conversion processes provide the dynamic links between the acoustic signals and auditory perception. At the phonological level, the features relating to contour, average pitch and voice quality not only represent the tones in more abstract terms but also reflect the native speakers' perceptual mechanisms which interpret the phonetic forms in accordance with their phonological structures.

I have provided data to demonstrate that the processes described are not arbitrary but they arise from the acoustic, articulatory and perceptual nature of Vietnamese tones as they are produced and perceived by native speakers in their speech.

Areas of Further Research and Possible Applications

The general approach and descriptive model proposed in this thesis can be useful in carrying on further research in areas which the present work has not been able to reach. One such area is the description of local, individual and contextual variation in more detail, where the various processes and rules proposed can be applied, tested and further refined or revised for better adequacy. Additional data in this area, especially those concerning the less known and less standard local dialects in remote regions might also contribute to a better understanding of the historical development of tone.

Another area is the application of the model to some form of computerized algorithms for automatic tone recognition (as suggested in Earle 1975), by elaborating on available data on Vietnamese tones and other relevant segmental and suprasegmental parameters. It would be possible, for example, to develop computer programs for automatic tone recognition by analyzing, step by step, the raw input of acoustic waves into physical phonetic parameters, normalizing them and converting them into phonetic and phonological tone features whose combination would be recognized as such or such tone.

I would predict that such programs could yield correct recognition of the tones in their standard forms, but further research would be necessary to handle variations and distortions occurring in connected speech.

BIBLIOGRAPHY

Abramson, A. S.

1972 Tonal experiments with whispered Thai. In A. Valdman (Ed) *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*. The Hague: Mouton. 1972: 31-44.

1975 The tones of Central Thai : some perceptual experiments. In Harris & Chamberlain (Eds) 1975: 1-16.

1976 Thai tones as a reference system. In Gething et al. (Eds) 1976: 1-12.

Anderson, S. R.

1978 Tone features. In Fromkin (Ed) 1978: 133-173.

Andreev, N. D. & Gordina, M. V.

1957 Sistem tonov Vjetnamskogo jazyka. *Vestnik Leningradskogo Gosudarstvennogo Universiteta*, 8: 132-148.

Bell, A. & Hooper, Joan B. (Eds)

1978 *Syllables and Segments*. Amsterdam: North Holland Publishing Company.

Berg, Jw. van den

1968 Mechanism of the larynx and the laryngeal vibrations. In Malmberg (Ed) 1968: 278-308.

Bolinger, D. (Ed)

1972 *Intonation*. Penguin Books.

Bùi Văn Nguyên *

1977 Thử tìm hiểu giọng nói Nghệ Tĩnh trong hệ thống giọng nói chung cả nước. *Ngôn ngữ*, 1977/4:34-41.

----

* Vietnamese authors are included alphabetically according to their surnames (first in order). For this purpose, no differentiation is made between the letters D and Đ (pronounced [z] and [d] respectively.)

Catford, J. C.

1964 Phonation types : the classification of some laryngeal components of speech production. In Abercrombie et al. (Eds) In Honour of Daniel Jones. London: Longmans.

1968 The articulatory possibilities of man. In Malmberg (Ed) 1968: 309-333.

1977 Fundamental Problems in Phonetics. Edinburg: Edinburg University Press.

Chang, Nien-Chuang T.

1958 Tones and intonation in the Chengtu dialect (Szechuan, China). In Bolinger (Ed) 1972: 391-413.

Chao, Y. R.

1930 A system of tone letters. Le Maître Phonétique, 30: 24-27.

Chen, M.

1970 Vowel length variation as a function of the voicing of the consonant environment. Phonetica, 22: 129-159.

Chomsky, N. and Halle, M.

1968 The Sound Pattern of English. New York: Harper & Row.

Chuang, C. K., Hiki, S., Sone, T., & Nimura, T.

1972 The acoustic features and perceptual cues of the four tones of standard colloquial Chinese. In Proceedings of the 7th International Congress of Acoustics (vol. 3) Budapest: Académial Kiado. 1972: 297-300.

Crystal, D.

1969 The intonational system of English. In Bolinger (Ed) 1972: 110-136.

Đái Xuân Ninh

1978 Hoạt động của từ tiếng Việt. Hanoi: NXB Khoa Học Xã Hội.

Delattre, P.

1962 Some factors of vowel duration and their cross-linguistic validity. Journal of the Acoustical Society of America, 14: 1141-1143.

Đoàn Thiện Thuật

1977 Ngữ âm tiếng Việt. Hanoi: NXB Đại Học và Trung Học Chuyên Nghiệp.

Dương Thanh Bình

1971 A Tagmemic Comparison of the Structure of English and

Vietnamese Sentences. The Hague: Mouton.

Earle, M. A.

1975 An Acoustic Phonetic Study of Northern Vietnamese Tones. Santa Barbara, California: Speech Communications Research Laboratory, Inc.

Emeneau, M. B.

1951 Studies in Vietnamese (Annamese) Grammar. Berkeley & Los Angeles: University of California Press.

Erickson, Donna

1975 Phonetic implications for an historical account of tonogenesis in Thai. In Harris & Chamberlain (Eds) 1975:100-111.

Ewan, W. G.

1976 Laryngeal Behaviour in Speech. PhD dissertation, University of California, Berkeley.

Ewan, W. G. & Krones, R.

1974 Measuring larynx movement using the thyroumbrometer. Journal of Phonetics, 2: 327-335.

Fischer-Jørgensen, Eli

1968 Les occlusives françaises et danoises d'un sujet bilingue. Word, 24: 112-153.

1972 p t k et b d g français en position vocalique accentuée. In A. Valdman (Ed) Papers in Linguistics and Phonetics to the Memory of Pierre Delattre. The Hague: Mouton. 1972: 143-200.

Flanagan, J.L.

1965 Recent studies in speech research at Bell Telephone Laboratories. Proceedings of the 5th International Congress on Acoustics, Liege.

Flanagan, J. L. & Landgraft, L.L.

1968 Self-oscillating source for vocal-tract synthesizers. IEEE Transactions: Audio and Electro Acoustics, AU-16,1: 57-64.

Fok Chan Yuen-Yuen

1974 A Perceptual Study of Tones in Cantonese. Hongkong: Centre of Asian Studies, University of Hongkong.

Fromkin, Victoria A. (Ed)

1978 Tone. A Linguistic Survey. New York: Academic Press.

Fudge, E. C.

1969 Syllables. Journal of Linguistics, 5: 253-286.

1973 (Ed) Phonology. Penguin Books.

Fujimura, O. and Lovins, Julie

1978 Syllables as concatenative phonetic units. In Bell and Hooper (Eds) 1978: 107-120.

Gandour, J. T.

1974 Consonant types and tone in Siamese. Journal of Phonetics, 2: 337-350.

1978 The perception of tone. In Fromkin (Ed) 1978: 41-75.

Gandour, J. T. and Harshman, R.A.

1978 Crosslanguage differences in tone perception : a multidimensional scaling investigation. Language and Speech, 21/1:1-33.

Gething, T.W., Harris, J.G. & Kullavanijaya, P. (Eds)

1976 Tai Linguistics in Honor of Fang-kuei Li. Bangkok: Chulalongkorn University Press.

Halle, M. and Stevens, K.

1971 A note on laryngeal features. Quarterly Progress Report, MIT, Research Laboratory of Electronics, 101: 198-213.

Han, Mieko S.

1967 Studies in the Phonology of Asian Languages V. Acoustic Features in the Manner Differentiation of Korean Stop Consonants. Los Angeles: University of Southern California, Acoustic Phonetics Research Laboratory.

1969 Studies in the Phonology of Asian Languages VIII. Vietnamese Tones. Los Angeles: University of Southern California, Acoustic Phonetics Research Laboratory.

Han, Mieko S. and Kim Kong-On

1972 Studies in the Phonology of Asian Languages X. Intertonal Influences in Two-Syllable Utterances of Vietnamese. Los Angeles: University of Southern California, Acoustic Phonetics Research Laboratory.

Harris, J.G. and Chamberlain, J. (Eds)

1975 Studies in Tai Linguistics. Bangkok : Central Institute of English Language.

Harris, J.G. and Noss, R.B. (Eds)

1972 Tai Phonetics and Phonology. Bangkok: Central Institute of English Language.

Haudricourt, A.G.

1954 De l'origine des tons en vietnamien. Journal Asiatique, 242: 68-82.

1961 Bipartition et tripartition des systèmes de tons dans quelques langues d'Extrême-Orient. Bulletin de la Société de linguistique de Paris, 56: 163-180.

1972 Two-way and three-way splitting of tonal systems in some Far-Eastern languages. In Harris and Noss (Eds) 1972:58-143.

Henderson, Eugénie J.A.

1961 Tonal exponents of pronominal concord in Southern Vietnamese. Indian Linguistics, 22: 86-97.

Hoàng Tuệ and Hoàng Minh

1975 Remarks on the Phonological Structure of Vietnamese. Vietnamese Studies, 40: 65-95.

Hồ Lê

1976 Vấn đề cấu tạo từ của tiếng Việt hiện đại. Hanoi : NXB Khoa Học Xã Hội.

Hollien H.

1974 On vocal registers. Journal of Phonetics, 2: 125-143.

Hombert, J.M.

1976 Perception of tones of bisyllabic nouns in Yoruba. Studies in African Linguistics, Supplement 6: 109-121.

1978 Consonant types, vowel quality, and tone. In Fromkin (Ed) 1978: 77-111.

Hombert, J.M. and Ladefoged, P.

1977 The effect of aspiration on the fundamental frequency of the following vowel. UCLA Working Papers in Phonetics, 36: 33-40.

Hombert, J.M., Ohala, J.J., and Ewan, W.G.

1979 Phonetic explanations for the development of tones. Language, 55/1: 37-58.

Hooper, Joan B.

1976 An Introduction to Natural Generative Phonology. New York: Academic Press.

House, A.S. and Fairbanks, G.

1953 The influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America, 25: 105-113.

Howie, J.M.

1976 Acoustical Studies of Mandarin Vowels and Tones. Cambridge: Cambridge University Press.

Hyman, L.M.

1973a (Ed) Consonant Types and Tone. Los Angeles: University of Southern California.

1973b The role of consonant types in natural tonal assimilations. In Hyman (Ed) 1973a: 151-179.

Ivić, P. and Lehiste, Ilse

1963 Prilozi ispitivanju fonetske i fonoloske prirode akcenatu u Savremenom Srpskohrvatskom knjizevnom jeziku. Zbornik za filogiju i lingvistiku, 6 (Noji Sad) : 33-73

Jeel, V.

1975 An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. Annual Report of the Institute of Phonetics, University of Copenhagen, 9: 191-211.

Jones, R.B. Jr and Huỳnh Sanh Thông

1957 Introduction to Spoken Vietnamese. Washington: American Council of Learned Societies.

Kagaya, R.

1974 A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. Journal of Phonetics, 2/2:161-180.

Kagaya, R. and Hirose, H.

1975 Fiberoptic, electromyographic and acoustic analyses of Hindi stop consonants. Annual Bulletin, Research Institute of Logopedics and Phoniatrics, 19: 27-46.

Kim, C.W.

1968 Review of Lieberman 1967. Language, 44: 830-842.

Kratochvil, P.

1968 Tone in Chinese. In Fudge (Ed) 1973: 342-353.

1971 An experiment in the perception of Peking dialect tones. Scandinavian Institute of Asian Studies monograph series, No 6.

Ladefoged, P.

1964 A Phonetic Study of West African Languages. Cambridge University Press.

1971 Preliminaries to Linguistic Phonetics. Chicago: The University of Chicago Press.

1975 A Course in Phonetics. New York: Harcourt Brace Jovanovich.

Laver, J.

1980 The Phonetic Description of Voice Quality. Cambridge University Press.

Lea, W.A.

1973 Segmental and suprasegmental influences on fundamental frequency contours. In Hyman (Ed) 1973a :15-70.

Leben, W.R.

1973 The role of tone in segmental phonology. In Hyman (Ed) 1973a: 115-149.

1978 The representation of tone. In Fromkin (Ed) 1978 : 177-219.

Lehiste, Ilse

1970 Suprasegmentals. Cambridge, Mass. : The MIT Press.

1978 The syllable as a structural unit in Estonian. In Bell and Hooper (Eds) 1978: 73-83.

Lehiste, Ilse and Peterson, G.E.

1961 Some basic considerations in the analysis of intonation. In Bolinger (Ed) 1972: 367-384.

Lê Văn Lý

1948 Le parler vietnamien. Paris: Editions Huong Anh.

Lieberman, P.

1970 A study of prosodic features. Haskins Laboratories Status Reports on Speech Research, SR-23: 179-208.

1977 Speech Physiology and Acoustic Phonetics. New York: MacMillan Publishing Co.

Löfqvist, A.

1975 Intrinsic and extrinsic Fo variations in Swedish tonal accents. Phonetica, 31: 228-247.

MacDonald, T.H.

1977 Statistics. Their Application to Human Activities. Pitman Australia.

Malmberg, B. (Ed)

1968 Manual of Phonetics. Amsterdam: North Holland Publishing Co (3rd Printing 1974).

Maspéro, H.

1912 Etudes sur la phonétique historique de la langue annamite. Les initiales. Bulletin de l'Ecole française d'Extrême Orient, 12/1: 1-127.

Matisoff, J.A.

1973 Tonogenesis in Southeast Asia. In Hyman (Ed) 1973a: 71-95.

Millar, J.B.

1974 Man-machine aspects of digital signal processing. ANU Computer Centre Technical Report, 45.

1978 An audio-visual waveform editing program. Proceedings of the Digital Equipment Computer Users Society, Canberra.

Miller, J.D.

1961 Word tone recognition in Vietnamese whispered speech. Word, 17/1: 11-15.

Mohr, B.

1968 Intrinsic fundamental frequency variation, II. Monthly Internal Memorandum. Phonetic Laboratory, University of California, Berkeley, June 1968: 23-32.

Nguyễn Đăng Liêm

1970 Vietnamese Pronunciation. Honolulu: University of Hawaii Press.

Nguyễn Đình Hoà

1967 Speak Vietnamese. Tokyo.

Nguyễn Hàm Dương

1962 Sistema tonov i spektry glasnyx vjetnamskogo jazyka. Kandidatskaja dissertacija. Moskva.

Nguyễn Phan Cảnh

1978 Bản chất cấu trúc âm tiết tính của ngônngữ. Dẫn luận vào một miêu tả không phân lập đối với âm vị học tiếng Việt. Ngôn ngữ, 2/1978: 5-18.

Nguyễn Tài Cẩn

1975 Ngữ pháp tiếng Việt. Hanoi: NXB Đại Học và Trung Học Chuyên Nghiệp.

Nguyễn Văn Tu

1976 Từ và vốn từ tiếng Việt hiện đại. Hanoi: NXB Đại Học và Trung Học Chuyên Nghiệp.

Ohala, J.J.

1972 How is pitch lowered? Journal of the Acoustical Society of America, 52: 124.

1973 The physiology of tone. In Hyman (Ed) 1973a : 1-14.

1978 Production of Tone. In Fromkin (Ed) 1978 : 5-39.

Petersen, N.R.

1976 Intrinsic fundamental frequency of Danish vowels. Annual Report of the Institute of Phonetics, University of Copenhagen, 1976: 1-27.

Peterson, G.E. and Barney, H.L.

1952 Control methods used in a study of vowels, Journal of the Acoustical Society of America, 24: 175-184.

Peterson, G.E. and Lehiste, Ilse

1960 Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32: 693-703.

Pike, K.L.

1948 Tone Languages. Ann Arbor: University of Michigan Press.

Rose, P.J.

1981 The Ningpo Dialects. Dissertation submitted for PhD degree, University of Cambridge.

Sampson, G.R.

1970 On the need for a phonological base. Language, 46: 586-626.

Sauvain, D.L.

1977 An Acoustical Analysis of Selected Yangchow Syllables. PhD dissertation, University of Cambridge.

Schuh, R.G.

1978 Tone rules. In Fromkin (Ed) 1978: 221-255.

Slis, I.H.

1966 A model for the distinction between voiceless and voiced consonants. IPO Annual Progress Report, 1: 40-44.

Thompson, L.C.

1959 Saigon phonemics. Language, 35/3: 454-476.

1965 A Vietnamese Grammar. Seatle : University of Washington Press.

1969 Some internal evidence for the history of Vietnamese tones. Bulletin of the Institute of History and Philosophy, Academia Sinica, 29/1: 415-423.

1976 Proto-Viet-Muong phonology. In Jenner, P.N., Thompson, L.C. and Starosta, S. (Eds) Austroasiatic Studies, Part II. Honolulu: The University of Hawaii Press.

Trần Hướng Mai, Aurélie

1969 Stress, Tones and Intonation in South Vietnamese. PhD Thesis, Australian National University.

Vũ Bá Hùng

1978 Thanh điệu - âm vị tuyến điệu của tiếng Việt. Ngôn ngữ, 1/1978: 13-23.

Vũ Thanh Phương

1978 Tone features in Vietnamese. Seminar paper given at A.N.U. Canberra.

1979 Tone perception in Vietnamese. Paper presented at the 11th Annual Conference of the Australian Linguistic Society, Newcastle, N.S.W.

1980 The dynamics of tone change: historical and phonetic evidence from Vietnamese. Paper presented at the 12th Annual Conference of the Australian Linguistic Society, Melbourne, Victoria.

1981 Phonetic properties of Vietnamese tones across dialects. To appear in Pacific Linguistics.

Vương Lộc

1975 Glimpses of the evolution of the Vietnamese language. Vietnamese Studies, 40: 9-30.

Wang, W. S.-Y.

1967 Phonological features of tone. International Journal of American Linguistics, 39: 93-105.

Yip, Moira Jean W.

1980 *The Tonal Phonology of Chinese*. Bloomington: Indiana University Linguistics Club.

Zee, E.

1978 The interaction of tone and vowel quality. *UCLA Working Papers in Phonetics*, 41: 53-67.