R E P O R T   R E S U M E S

ED 011 106 AL 000 051

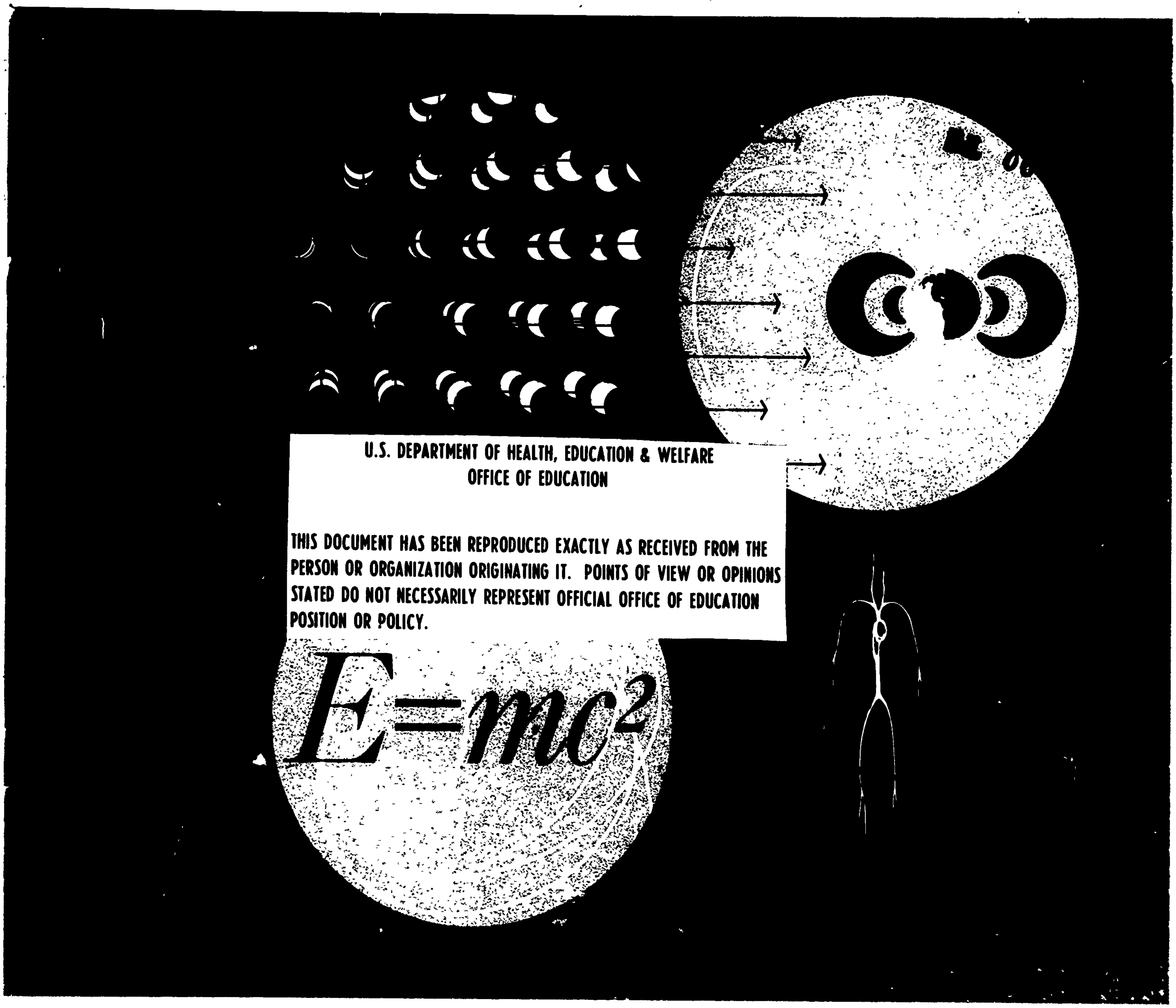
INSTRUMENTAL INVESTIGATION OF SINGLE AND PAIRED MANDARIN TONEMES.

BY- DREHER, JOHN J.   LEE, PAO-CH'EN
DOUGLAS AIRCRAFT CO., HUNTINGTON BEACH, CALIF.

REPORT NUMBER DARL-RES-COMMUNICATION-13   PUB DATE   SEP 66
REPORT NUMBER DOUGLAS PAPER-4156
EDRS PRICE MF-$0.18 HC-$2.72   68P.

DESCRIPTORS- CHINESE, *MANDARIN CHINESE, *PHONETIC ANALYSIS, PHONETICS, MEASUREMENT, INSTRUMENTATION, HUNTINGTON BEACH

THE AIM OF THIS STUDY WAS TO IDENTIFY AND SET TONE-REGISTER BOUNDARIES, AND TO DESCRIBE TONAL CONTOURS OF SINGLE WORDS AND COUPLETS IN TERMS OF WHAT GOOD MANDARIN CHINESE SPEAKERS ACTUALLY DO. THE STUDY WAS BASED ON ISOLATED AND COUPLED WORDS OF HIGH LINGUISTIC FREQUENCY, WHICH WERE RECORDED BY 18 PROFESSORS OF MANDARIN CHINESE ON NARROW-BAND SONOGRAMS. FOR STANDARDIZATION AND CONVENIENCE, INDIVIDUAL DIFFERENCES WERE REMOVED BY REPRESENTING TONE BEHAVIOR IN TERMS OF CONTOUROIDS. THE RESULT WAS A MATHEMATICALLY AND LINGUISTICALLY COMPARABLE SET OF CONTOURS THAT CAN BE DISPLAYED TO EXAMINE THE PERFORMANCE OF DIFFERENT SPEAKERS USING THE SAME SET OF GROUND RULES. CATEGORIES INCLUDED TONES ALONE, PRENEUTRAL POSITION, NEUTRAL BEHAVIOR, AND THE EFFECTS OF PRE- AND POST-TONE POSITION ON COUPLETS. (IT)



INSTRUMENTAL INVESTIGATION OF SINGLE AND PAIRED MANDARIN TONEMES

J. J. DREHER
P.C. LEE

SEPTEMBER 1966

DOUGLAS CORPORATE OFFICES · ADVANCED RESEARCH LABORATORY

DOUGLAS PAPER 4156

DOUGLAS CORPORATE OFFICES

INSTRUMENTAL INVESTIGATION OF SINGLE AND PAIRED MANDARIN TONEMES

by
JOHN J. DREHER
Advanced Research Laboratory
Douglas Aircraft Company, Inc.
Huntington Beach, California

PAO-CH'EN LEE
Defense Language Institute
Monterey, California

SEPTEMBER 1966

ADVANCED RESEARCH LABORATORY · 5251 BOLSA AVENUE · HUNTINGTON BEACH, CALIFORNIA

CREDIT

Work presented herein was conducted by the Douglas Advanced Research Laboratory under company-sponsored Research and Development funds.

FOREWORD

Ever since man has talked, he has attempted to analyze his language. Motivated variously by superstition, pedagogy, commerce, or cultural continuity, the Chinese analyst has been making subjective constructs of his language since the turn of the millenium.

With the advent of modern technology, new dimensions of this traditional study are needed to describe and identify languages for computer handling and other types of machine processing. Acknowledging the eventual political and possible commercial importance of the Chinese language, as well as its cultural value, this study represents the first of a series of linguistic investigations in the Douglas Advanced Research Laboratory.

PRECEDING PAGE BLANK-NOT FILMED

ACKNOWLEDGMENTS

Thanks are due to Colonel Richard J. Long, Commandant of the Defense Language Institute, Monterey, California, to Professor Tsutoumu P. Tekawa, Director of the Far East Division, Defense Language Institute, and to their staff for collaboration in this series.

Grateful acknowledgment is given to Mrs. Elaine Young, Research Associate at Douglas Advanced Research Laboratory, for her indispensable aid and assistance in this study.

The authors also wish to express appreciation for the generous help extended by Professors William S. H. Pang, Victor Wen, Patrick Yang, Ella Chow, Yat-shek Sun, Chi Fu, Shu-min Mau, Poo-li Chu, Pai-fang Lee, Albert Fei, Rio Liang, Tsi-ying Chang, Meili Chen, Kuang-chin Cheng, Chih-kang Wu, Tzch-nan Chang, and Lindberg Kao.

PRECEDING PAGE BLANK- NOT FILMED

ABSTRACT

The commonly accepted five registers of Mandarin are verified and shown each to be approximately 36 cps wide, with a standard deviation of 1.29 cps on a normalized scale. The 68-percent normal population limits are delineated for the emphatic tone forms, as well as new 5-register number notations for tones actually produced in isolation and in couplets. A complete permutation of toneme environment contouroids is included.

The study was based on isolated and coupled words of high linguistic frequency, which were recorded by eighteen professors of the Mandarin language. The pitch contours and protensities of their utterances, displayed as normalized frequency-change ratios and durations, were grouped according to tonemic context and compared for experimental consistency with postulated theoretical behavior. For standardization and convenience, individual differences were removed by representing tone behavior in terms of contouroids (simplified contours). Categories include tones alone, pre-neutral position, neutral behavior, and the effects of pre- and post-tone position on couplets.

EXECUTIVE SUMMARY

This paper uses certain physical and mathematical methods that may not be orthodox for the descriptive linguist. This summary is therefore supplied to answer some questions that might otherwise plague the reader.

The Mandarin dialect uses tonal inflections of the voice to help distinguish different word meanings. Identification of these tonal contours has long been postulated to be a function of certain definite pitch registers of the voice. This is both a logical and useful concept. One basic drawback, however, is that the boundaries of these registers are fuzzy and no standard method exists for comparing what people supposedly do with what they actually do. With this "floating index," so to speak, no quantitative description useful for eventual machine handling methods has been available.

The aim of this study was to identify and set tone-register boundaries, and to describe tonal contours of single words and couplets in terms of what good Mandarin speakers actually do. Hopefully, the methods and resulting parameters would aid further progress toward eventual mechanical translation.

Eighteen speakers, all professors at the Defense Language Institute, Monterey, California, recorded some common words and combinations for analysis. These words were measured in several ways. Some 5000 actual and transformed data points allowed us to describe in a statistically and psychophysically reliable fashion what the Mandarin contours look like, what their probable limits are, how they relate

to the registers, and what determines the dimensions of the registers themselves.

First we made narrow-band sonagrams. Next, we calculated the movement of the voice fundamental (which largely correlates with pitch perception) at certain points within the utterance. We then made a simplification: an upward glide, for instance, was considered describable by a straight line connecting the upper and lower frequency limits. Since the listener is really concerned with direction and the starting and ending registers rather than with contour, we felt that this procedure was permissible and pertinent for purposes of comparison.

This procedure generated a libretto of what we call "contouroids," or contour-like figures, leaving out the nonessential distractions of individual voice quirks.

As the tonal linguist well knows, as long as the tonemes of a language consistently retain their own relationship to each other, it does not matter whether a basso or soprano does the speaking. More simply, the absolute voice register itself does not contribute to meaning in Mandarin. This fact then allows us to bring further order out of the chaos of individual observation by using certain normalizing techniques.

Basically, this is what is involved: a fourth tone, for instance, as spoken by different people, will show various starting pitches, varying degrees of drop, and different durations. How, then, may one compare them? The procedures we have used herein enable us to make certain compensations. These preserve all significant features, yet allow meaningful comparison.

The first step translates all starting pitches to the same point (individual voice register is nonsignificant); the second calculates an average duration (to get a typical behavior); the third step then calculates a pitch change equivalent to the original contour in terms of how the ear perceives pitch difference. The result is a mathematically and linguistically comparable set of contours that may be displayed to examine the performance of different speakers using the same set of ground rules.

This procedure was followed with all the isolated and paired tones produced by the eighteen informants in this study.

Here are some relationships that came to light:

- Mandarin does show five definable levels of pitch register as postulated by Chao.[3] In terms of our experimental voices and their normalization, the registers are 35.7 cps wide, with a standard deviation of 1.29 cps.

- In terms of these registers, an overall description of the four Mandarin tones spoken in isolation takes the following form:

  Tone 1: 54
  Tone 2: 24
  Tone 3: 213
  Tone 4: 51

- When paired with each other, the four tones subtly change their contours, as indicated in Table 3, page 39.

- Neutral or zero tones are not simply points on the frequency scale, but very short, fully contoured tonemes, described as follows:

  Post-1 tone: 41
  Post-2 tone: 31
  Post-3 tone: 23
  Post-4 tone: 21

- In paired tones (couplets), the tone following is longer than that preceding and shows less influence on its shape from the pairing process.

- The emphatic (isolated) tone is longer than the same tone when paired with another.

- Only two registers exist for the start of all four tones: 4 and 1 originate in the same register; 2 and 3 also originate together.

- Both tones 1 and 4 start lower when used in combination than when standing alone.

- Tone 3 shows its expected pre-3 sandhi form as tone 2; when preceding tone 1, its behavior may show either a fall-rise or a straight fall. When preceding tone 4, it may show either a fall-rise or a straight rise (see Figures 15 and 16, pages 33 and 34).

- When tone 3 occurs after tone 1, it may show either a fall-rise or a straight rise (Figure 17, page 35).

- Tone 4 falls more sharply when in a following than when in a preceding position.

We have included diagrams of the tone forms and their values, both as single utterances and as members of a pair of tones.

The next study will be an application of this analysis method to three-word sequences, and subsequent efforts will be aimed at exploration of the other main dialects.

TABLE OF CONTENTS

Section		Page
1	INTRODUCTION	1
2	EXPERIMENTAL PROCEDURE	7
	Aims	8
	Methods	8
	Contouroid Concept	10
3	EXPERIMENTAL FINDINGS	15
	Emphatic Forms	15
	Establishment of Registers	18
	Tones and Their Neutrals	23
	Sandhi Effects on Tonemes	28
	Duration of Tonemes	40
	Remarks	47
4	NOTES	49

PRECEDING PAGE BLANK- NOT FILMED

LIST OF ILLUSTRATIONS

Figure		Page
1	Five tones of Peiping-Mandarin (after Liu Fu, 1924)	5
2	Toneme analysis instrumentation	9
3	Mechanics of tone normalization	14
4	Behavior of isolated tones	17
5	Mandarin tone registers	19
6	Emphatic tones (±1-sigma limits)	22
7	Neutral tones (±1-sigma limits)	24
8	Contours of tones in preneutral position	25
9	Neutral tones	26
10	Tones plus neutrals	27
11	Tone 1 contouroids -- isolation, pre-neutral, 1, 2, 3, 4	29
12	Tone 1 contouroids -- post 1, 2, 3, 4	30
13	Tone 2 contouroids -- isolation, pre-neutral, 1, 2, 3, 4	31
14	Tone 2 contouroids -- post 1, 2, 3, 4	32
15	Tone 3 contouroids -- isolation, pre-neutral	33
16	Tone 3 contouroids -- pre 1, 2, 3	34
17	Tone 3 contouroids -- pre 4	35

LIST OF ILLUSTRATIONS (Continued)

Figure		Page
18	Tone 3 contouroids -- post 1, 2, 3, 4	36
19	Tone 4 contouroids -- isolation, pre-neutral, 1, 2, 3, 4	37
20	Tone 4 contouroids -- post 1, 2, 3, 4	38

LIST OF TABLES

Number		Page
1	Tone Analysis Categories	15
2	Normalized Measure Register Intervals	20
3	Contours of Tones in Couplet Combination	39
4	Duration of Mandarin Tonemes in Isolation and in Couplets	41
5	Effect of Couplet Environment on Duration of First Tones	42
6	Effect of Couplet Environment on Duration of Second Tones	43
7	Effect of Couplet Environment on Duration of Third Tones	44
8	Effect of Couplet Environment on Duration of Fourth Tones	45
9	Effect of Couplet Environment on Duration of Neutral Tones	46
10	Durations of Emphatic Tones	46

Section 1

INTRODUCTION

Modern Chinese, with its some 50-odd identifiable dialects, constitutes the largest language family in the world. It is a "tone contour" language, the words of which are invested with tonal glides, in contradistinction to "tone register" languages such as Ibo, which utilize various level voice registers. Chinese today is represented by five major dialects. The most important of these is Mandarin, which uses four lexical tones.

Historically speaking, both the phonetic and tonal structure of Mandarin are relatively new, and in its streamlined present form it acts as a lingua franca in China. As such, it promises to increase in both military and commercial importance.

Since early times, Chinese linguists have concerned themselves with descriptions of their language, often attaching mystical powers to certain sound combinations and incantations. The purely analytical approach has also been followed, some of the earliest works being written in Sanskrit, and, like the commentaries of Egypt, products of the leisure-class scholar.

During the medieval period, particular attention was directed toward the tones themselves. Chou Yung, a scholar of the fifth century, A.D., was the first to write a book about the so-called "four tones" entitled Szsheng Chyeyun (Pronunciation of the Four Tones). Later, Shen Ywe, a contemporary scholar of Chou Yung, wrote his Szshengpu

(Treatise on Four Tones), and was quite pleased with his discovery and interpretation of the four tones (which, according to him, earlier scholars were unaware of. Emperor Wudi, evidently displeased with Shen Ywe's arrogance, asked Chou Yung what these four tones were that Shen Ywe bragged so much about. Chou Yung gave the emperor a four-word example -- tyandz shengjê,* "Your Majesty is sage and wise," -- which represented the original four tones, and in their consecutive order, too: ping, shang, chyu, and ru (even, low, going, and entering). Later, when ping was split into yinping and yangping, and ru was abolished and distributed among the other tones, they became the present four tones in Mandarin: level, rising, low, and falling. Fortunately, je was distributed to the rising-tone group, so the quotation from Chou Yung still represents the four tones in present Mandarin -- tyāndž shèngjé (level, low, falling, and rising), but unfortunately not in consecutive order.

By certain literary methods of sound reconstruction, the tonal system of the Chou dynasty, 909-255 B.C., as represented by the Book of Odes, shows that three tones were used in the Chinese of that period. In the five main dialects spoken in present-day China, there are four tones in Mandarin, eight or nine in Cantonese, seven or eight in Min (Fukien), seven or eight in Wu (Soochow), and six in Hakka.

*An arbitrary tone mark representing the rusheng (entering tone), now obsolete in the Mandarin dialect (the national language).

The exact number of tones and their contour variations have obviously not been matters of universal agreement among scholars. The traditional methods of the descriptive linguist have yielded completely workable systems for teaching the language which have been used for centuries. What makes these systems workable, however, is man himself as the student, and his as-yet undefined heuristic processes. Machine handling of languages, on the other hand, makes mandatory the establishment of statistically expectable parameters and variations. The problem obviously cannot be simply or neatly structured in view of the many influences simultaneously involved in tone variation:

Lexical tone contours

Sandhi and accommodation

Stress and duration influences

Individual speaker variation

Emotional contour overlays

Dialectical influence on toneme contour

Possible arbitrary tone sandhi

Contour change by word elision

All the foregoing influences may singly or severally act to alter a toneme contour drastically when spoken by a particular informant. Because of the necessities of teaching and analyzing whole dialects, however, it is usually the procedure of the descriptive linguist to hew out broad generalities that can be reliably observed and to set these down as basic rules of the language or dialect. Although the available scientific literature has been understandably sparse in instrumental examination of

Chinese tones, two studies are worthy of comment. Liu Fu,[1] in a kymograph investigation, remarkably sophisticated in technique for the era of 1924, furnished tracings of the four tones by one informant for Peiping-Mandarin, as well as several other dialects by individual residents in France at the time. Due to the limitations of his instrument, as well as the fact that his Peiping-Mandarin informant elected to pronounce the fourth tone test word with a falling-rising inflection, the four-contour ensemble disagrees somewhat with both our popular conceptions and the contours shown in this study (see Figure 1).

Liu Fu's Tone 1 (Figure 1) is shown with a slightly rising contour, although it could possibly be interpreted as "level" (according to the register concept), depending upon the register boundaries set up by that particular informant. Tone 3, shown as a low-starting rise, disagrees with the present study's unanimous 18-voice portrayal as a fall-rise contour. Liu's careful mathematical treatment of his data, however, probably salvaged most of the information from the kymograph.

A considerably advanced technique has been applied by Wang and Li,[2] using electronic pitch-circuitry. Done in 1964, and in the context of machine-recognition cues, this study is addressed to a problem considerably wider than that of tones alone, and hence does not attempt to set norms. Consequently it adopts the sometimes commonly accepted descriptive contours 55, 35, 315, and 51.

Following is a description of an experimental quantification of Mandarin emphatic and couplet contours.

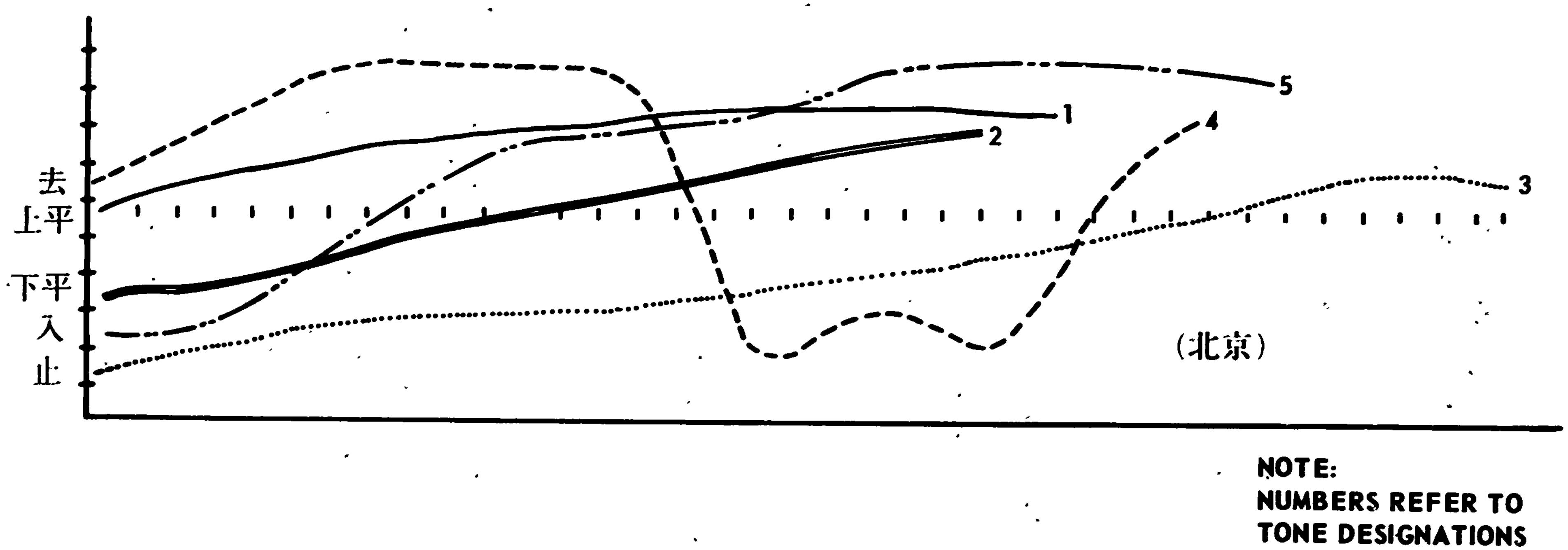


Figure 1. Five Tones of Peiping-Mandarin (After Liu Fu, 1924)

Section 2

EXPERIMENTAL PROCEDURE

It is generally conceded that the basic structure of Mandarin involves four basic tone contours, and a fifth, consisting of a "neutral" or "zero" tone, employed according to the dictates of morpheme combination. These contours have been variously represented by numbers, diacritical marks, quasi-musical staff notations, and verbal descriptions.[3]

To simplify a first approach, three environments of production were investigated to get some basic measurements of "pitch" contour and morpheme duration. These environments consisted of the four tones in isolation, the four tones followed by a neutral, and the four tones paired with themselves and each of the remaining three 2-morpheme sequences. In the number notation system, the combinations would read as follows:

Isolation	Tone + Neutral	Tone Pairs	
1	1 - 0	1 - 1	3 - 1
2	2 - 0	1 - 2	3 - 2
3	3 - 0	1 - 3	3 - 3
4	4 - 0	1 - 4	3 - 4
		2 - 1	4 - 1
		2 - 2	4 - 2
		2 - 3	4 - 3
		2 - 4	4 - 4

AIMS

The experimental aims in relation to these categories are as follows:

(Isolated Tones):	To determine the extent of, adherence to, or variation from the posited contour descriptions[3] and to note these as a duration and physical excursion of the voice fundamental.
(Tone + Neutral):	To fix the duration, register, and contour (if any) of the neutral tone as a function of its pairing with a given tone.
(Tone Pairs):	To examine duration and tone sandhi effects resulting from paired juxtaposition of the four tonemes with each other.

Eighteen speakers of Mandarin, instructors at the Defense Language Institute, Monterey, California, recorded words and phrases[4] embodying the tonal environments noted above. Recordings were made with an Electro-Voice, Model 665 (dynamic cardoid) microphone and an Ampex 601 recording system in the Far East Division recording studio at the Institute. Prior to the taping, each informant was instructed in the recording mechanics and desired unemotional structure of the utterances.[5] All recording and analyses were performed at 7-1/2-ips tape speed.

METHODS

All tapes were analyzed on the equipment shown schematically in Figure 2. Recording the words on a 50-5000 cps scale,

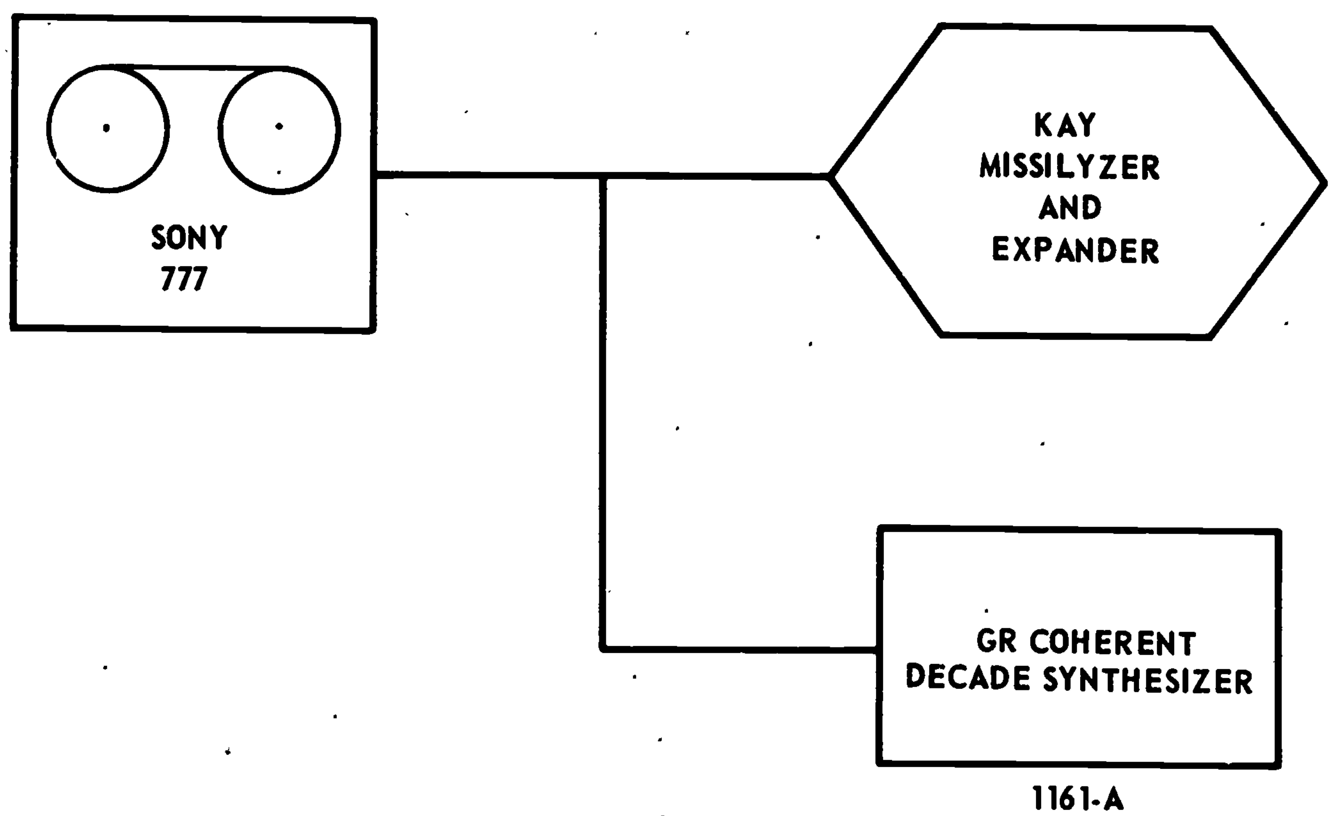


Figure 2. Toneme Analysis Instrumentation

and using a 2x narrow-band expanded scale spectrogram constituted a satisfactory compromise among the machine variables. Frequency calibrations were inserted on spectrograms by use of the Coherent Decade Frequency Synthesizer in 10-cps steps. All data points relating to fundamental frequency of the voice were calculated from readable harmonics.[6]

CONTOUROID CONCEPT

Display of these signals was approached from several points of view, each one helping to answer a different type of question.

Unless one is a clinician interested primarily in individual differences, retention of these characteristics acts only as noise, obscuring the fundamental relationships which are generally the invariances sought by the linguist. Therefore, the substitution of a "contouroid" or contourlike quantity[7] for the voice trace actually produced was a necessary first step. For statistical handling, these idealized contours seem to perform quite adequately, and embody the tonal behavior most closely allied to other sensory perception of "what is happening" during tonal excursions.

Following the recognized conventional linguistic description, both the lexical and sandhi tone contours of Mandarin are assumed to have a consistent gross continuity. This is to say, a first tone is described as "high-level," and the dominant linguistic impression supposedly exhibited by this sound is that of a high-register constant pitch. In point of fact, however, inspection of instrumental voice traces of first tones shows that their presumed constant frequency may be subject to several qualifications -- a sort of tremolo in some instances, a certain non-systematic internal contouring of the tone in others, and so on. Since, however, the overall contrastive effect is more "high-level" than are any of the other three tonal contours, the concept serves quite effectively as a description for that particular tone.

Focusing, then, on the significant function of the particular tones, the graphical representations for tones 1, 2, and 4 are assumed to be straight-line functions of average slope, while the tone 3 is represented as two straight lines connected to a minimum point, disregarding the different variations of actual concavity or convexity of the contours observed in the instrumental traces.

To derive typical and comparable representations of the tones as produced by the eighteen subjects, a normalization process was used, as explained below.

Since register displacement of the tone contours does not interfere with their relative functions or identity,[8] the fundamental frequencies were first transformed at the point of inception to a common point of origin.

The limited number of informants, as well as the differences between male and female voices, make any absolute value of the inception pitch a meaningless quantity. Its value, however, as a relative pitch starting-point, as compared with other tones produced by the same group of informants, is obvious. Therefore, the inception pitch was established numerically as the arithmetic average of the several starting pitches of the informants. This was done both for isolated words and word combinations.

Now that the common inception pitch had been established, a rationale for the most meaningful pitch-slope measure was needed.

In the frequency ranges of interest here -- namely, in the two octaves from 50-200 cps -- the interpretation of pitch interval is best expressed as a ratio of frequency change to the fundamental, rather than as an absolute frequency difference. This is to say, the ear matches a 100-105 cycle interval with a 200-210 cps interval, not a 200-205 cps interval. Consequently, the amount of pitch change depends not upon the actual physical frequency variation, but upon the relation of that quantity to the original starting frequency. Using such a criterion, the ear is able to function quite well with tone languages, regardless of the speaker's voice register. This is, of course, the psychophysical basis for Pike's linguistic observations.[8]

Where the pitch interval change is also associated with duration, that is, where the frequency difference is spread over some intervening time, rather than existing as two discrete occurrences as tones, a measure of that duration is also required. Thus the gestalt of a tonally inflected word depends upon its length as well as its change in frequency. The tone change is represented in this study as distributed over the whole length of the syllable. To establish a common value for duration for words within any class, this duration was expressed as an arithmetic mean quantity.

The common inception frequency F and durations M having been set for the class of tones, the final step in the normalization was the adjustment of each individual utterance of frequency change Δf in terms of these calculated parameters. Since the relation of pitch change $\left(\text{frequency difference related to inception frequency, } \frac{\Delta f}{f_o}\right)$ to tone duration ℓ must be equivalent to the normalized pitch change $\frac{\partial f}{F}$ in respect to the normalized duration M, the relation for calculating ∂f may be stated as follows:

$$\frac{\frac{\Delta f}{f_o}}{\ell} \approx \frac{\frac{\partial f}{F}}{M}$$

and the normalized frequency change for a single utterance may be expressed

$$\partial f = FM\left(\Delta f \ell^{-1} f_o^{-1}\right)$$

The class mean value ∂f now stands for the typical frequency change value for the particular tone in its stated context. Figure 3 illustrates this operation for a typical group of second tones.

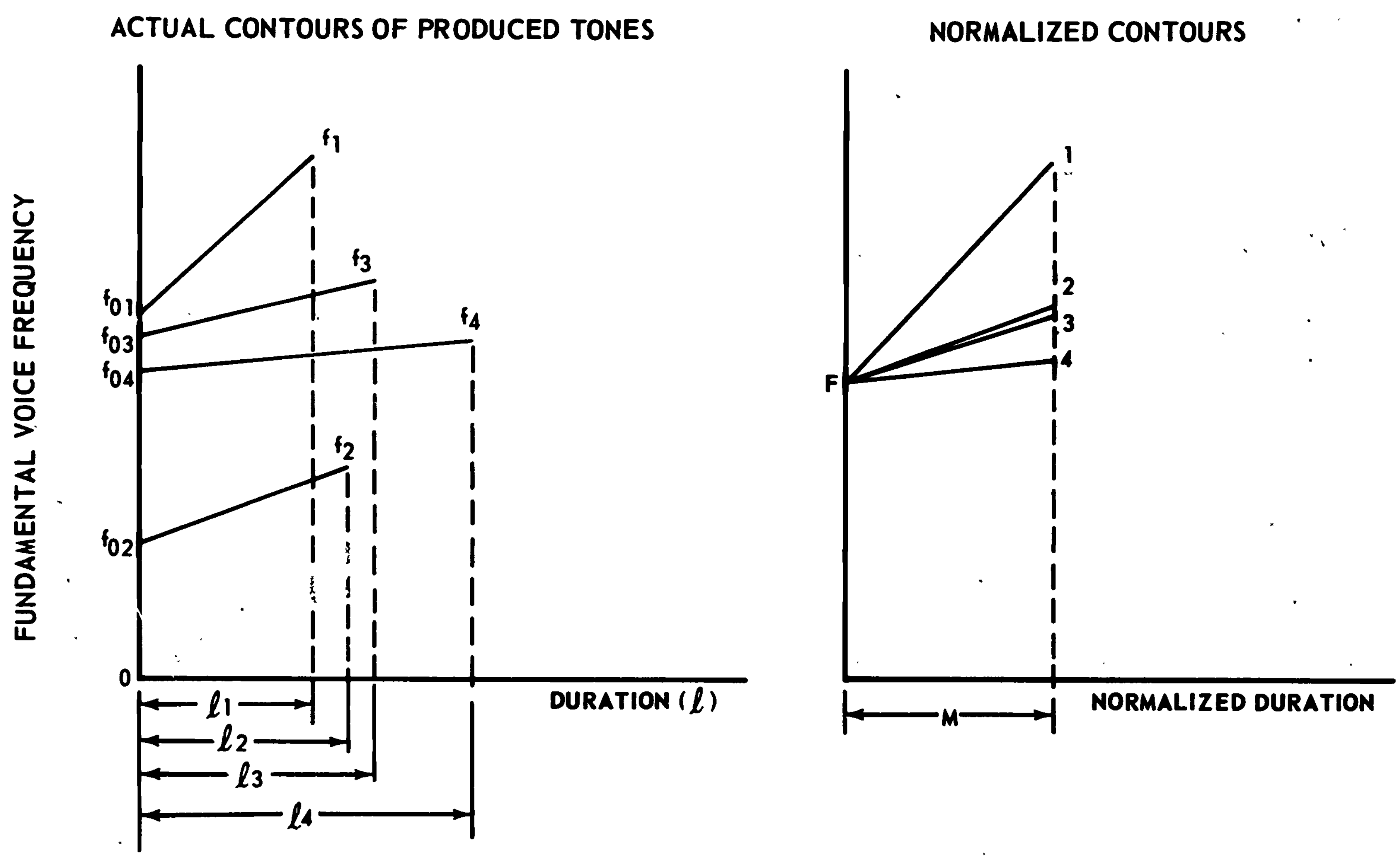


Figure 3. Mechanics of Tone Normalization

Section 3

EXPERIMENTAL FINDINGS

Because the reader will be interested in various levels of abstraction of tonal behavior, the analyses follow the model of Table 1.

EMPHATIC FORMS

It is the "emphatic" form, the contour produced when a word is stressed or pronounced in isolation, that comes to mind when a "typical" toneme is referred to. Consequently, the display and comparison of these tonemes becomes the starting point from which to compare the sandhi forms.

Table 1

TONE ANALYSIS CATEGORIES

Tone	Condition	Purpose
1,2,3,4	Isolation	Emphatic contours
Neutral	Post 1,2,3,4	Effects on starting pitch; neutral contour; agreement with theory
1,2,3,4	Pre-neutral	Pre-neutral behavior
1,2,3,4	Pre-, post 1,2,3,4	Sandhi effects, mean slopes, starting pitches

Figure 4, which shows the four basic Mandarin tones plotted on a time-frequency scale, might be thought of as quantitative equivalents of the Yale tone marks. There are several aspects of immediate interest in Figure 4. One is the fact that only two registers are used for starting points, and another is the shorter durations used with those tones starting in the upper register (1 and 4).

It is the relationship set up among these emphatic forms that leads to the establishment of the number and size of registers employed by the language. Their description, considered in the next section, is based upon some of the following statistically supportable observations:

- Tones 1 and 4 begin at statistically the same level, which is higher than the beginning level of Tones 2 and 3.
- During production, Tone 1 falls with a slope of 0.07, or $\frac{\partial f}{M}$, where ∂f is in cps and M is in milliseconds.
- Tones 3 and 2 start at statistically the same level.
- Tones 1 and 2 end at statistically the same level.
- The endpoint of Tone 3 stands alone in the approximate center of the complete tonal range.
- The endpoint of Tone 4 and the inflection point of Tone 3 are lower than the beginnings of Tones 2 and 4.
- Durations of 2 and 3 are statistically the same.
- Durations of 1 and 4 are statistically different.
- Durations of 1 and 2 are statistically different.
- Durations of 1 and 3 are statistically different.
- Durations of 2 and 4 are statistically different.
- Durations of 3 and 4 are statistically different.

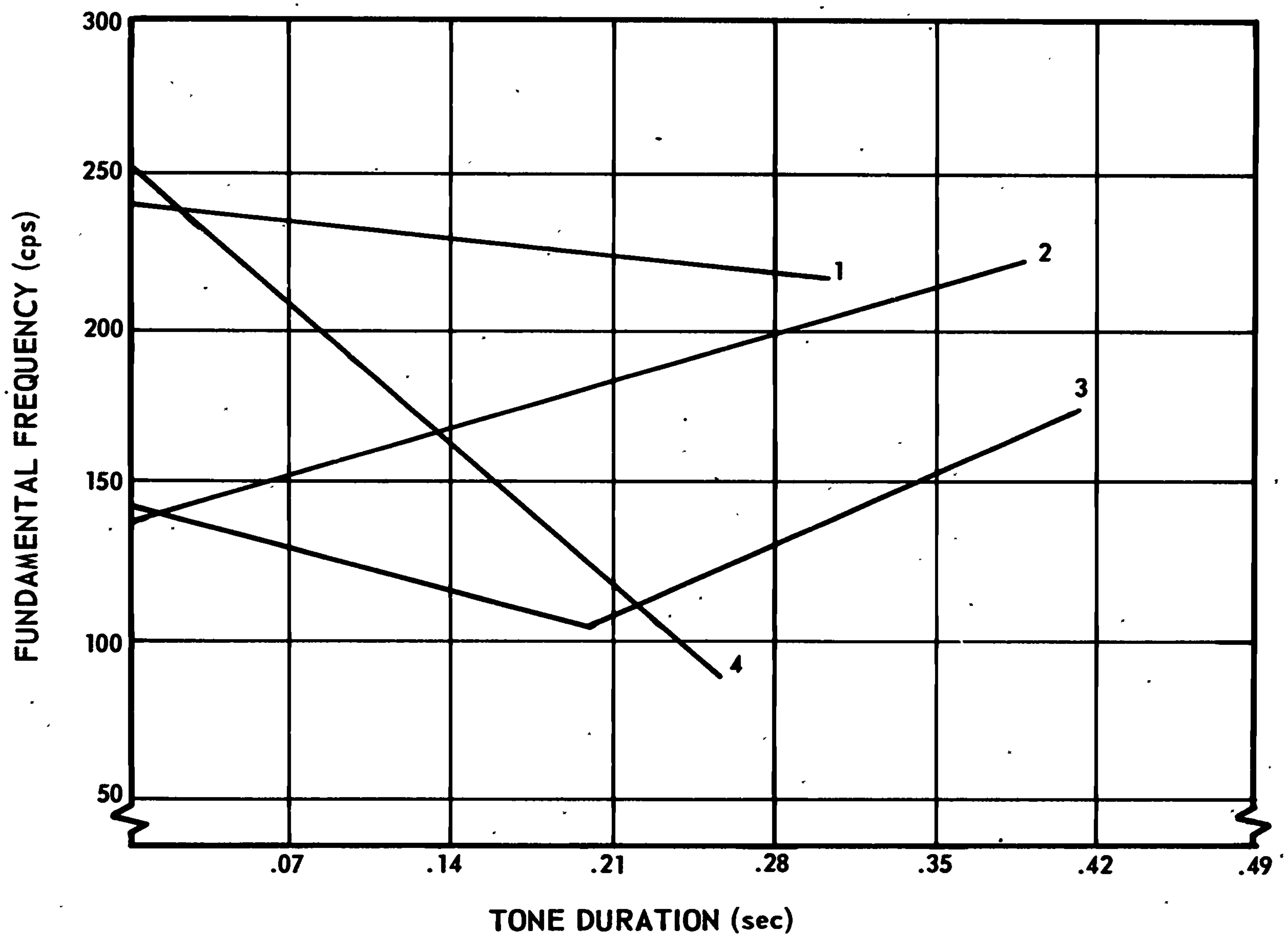


Figure 4. Behavior of Isolated Tones

- Tone 4 falls with a slope of -0.64.
- Tone 2 rises with a slope of 0.21.
- Tone 3 falls and rises with respective slopes of -0.19 and 0.33.

ESTABLISHMENT OF REGISTERS

In defining "register" we are here referring to frequency (or pitch) regions which statistically and consistently contain events and which are mutually exclusive of other events. On the basis of the observations made above of emphatic beginning, end, and inflection points, with their statistical equivalences and differences, we may define five distinct registers for Mandarin, located as follows:

Register I	Endpoint of 4, inflection point of 3
Register II	Beginning of 2 and 3
Register III	Endpoint of 3
Register IV	Endpoints of 1 and 2
Register V	Beginnings of 1 and 4

According to these groupings, and from Figure 5, the intervals appear to be of almost equal width, with bottom and top registers open-ended.

The method of calculation made the assumption that the range could be evenly split between the bottom value of one class and the top value of the next lower class. This, as a matter of fact, is the only permissible assumption without additional evidence on the psychophysical intervals. The net result is an estimate of the boundary limits of the

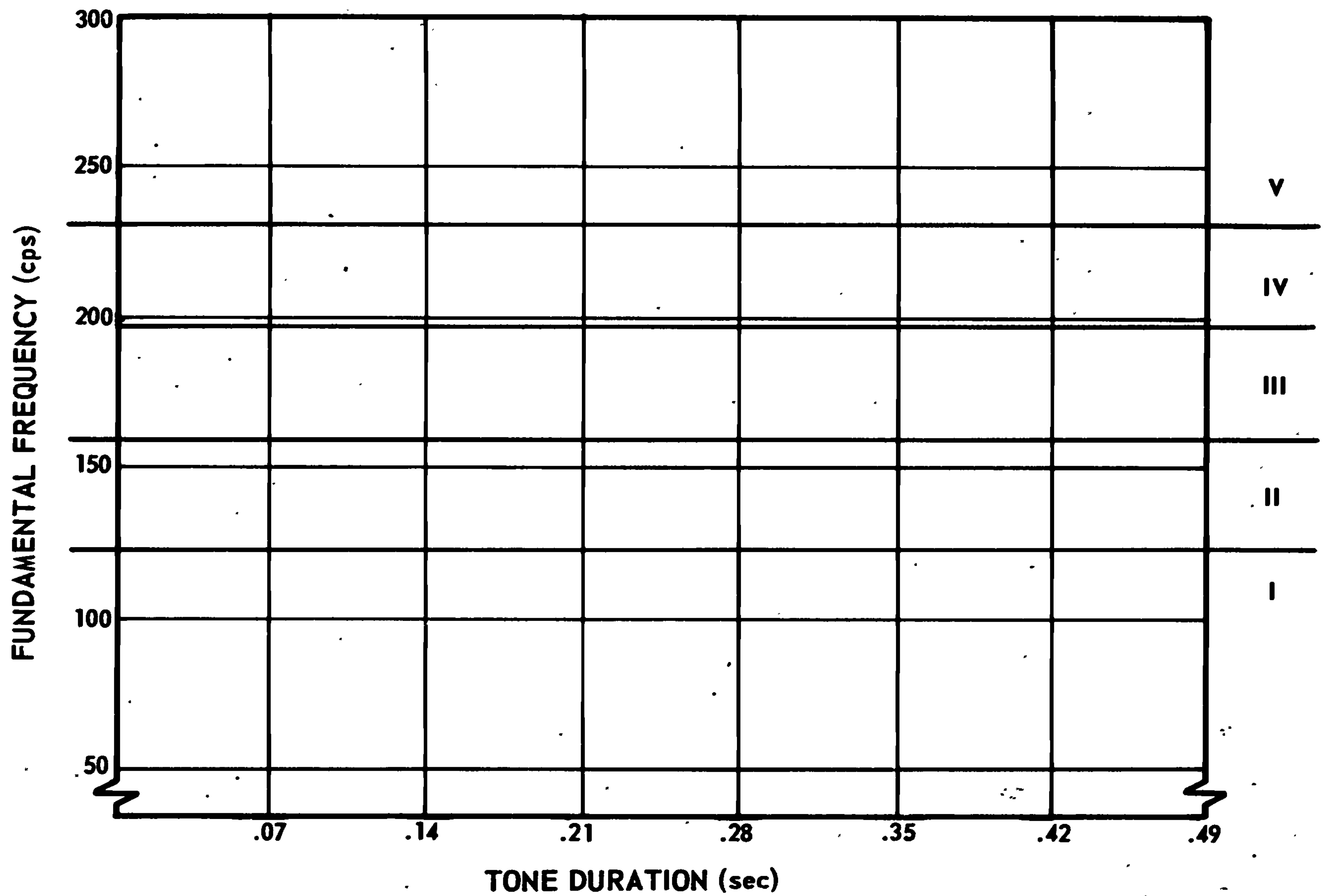


Figure 5. Mandarin Tone Registers

5-register notation system, as well as a template by which to compare sandhi effects.

Considered in terms of this study, with the normalized measures used, the register intervals may be expressed as in Table 2.

Table 2

NORMALIZED MEASURE REGISTER INTERVALS

Register	Limits		Width (cps)
	Lower (cps)	Upper (cps)	
1	-----	122.0	17.8 (half width)
2	122	159.3	37.3
3	159.3	196.2	36.9
4	196.2	230.7	34.5
5	230.7	-----	17.0 (half width)

If we wish to consider the open-ended bottom and top registers as approximately equally divided by the experimental points, we arrive at a mean estimate of each interval as being 35.7 cycles wide, with a standard deviation $\sigma = 1.29$ cps.

Absolutes in voice pitches or duration are not implied by Figures 4 and 5, nor by Table 2. We are studying the relationships of these two aspects as embodied by the linguistic behavior of admittedly good practitioners of the dialect, and it is, to be completely precise, relationships only that are of concern to linguistic theory (as distinct from "language"). Conclusions made in this domain are as reliable as the experimental controls exercised.

The existence of such a register template now allows us meaningfully to relate any observed differences in tones as a function of their environment.

Using these emphatic registers as a reference, we can designate the four basic emphatic contours as follows:

1: 54

2: 24

3: 213

4: 51

Such a designation is quite close to the presently conceived notion, differing only on the first and second contours (54 instead of 55, and 24 instead of 35).

Because statistics are descriptive, not prescriptive, the pitch profiles shown above are not necessarily recommended for teaching techniques. These new contour designations are merely meant to show what speakers actually do in terms of our standard notation, not what they should strive to do. Certainly if the first tone is defined as "high-level," an attempt at 55 should produce better overall performance than an attempt at 54, with the attendant variations that would result.

Figure 6 shows the four tones plotted with ±1-sigma "fans." These fans delimit the variation possible with 68 percent of all speakers of the type who served as

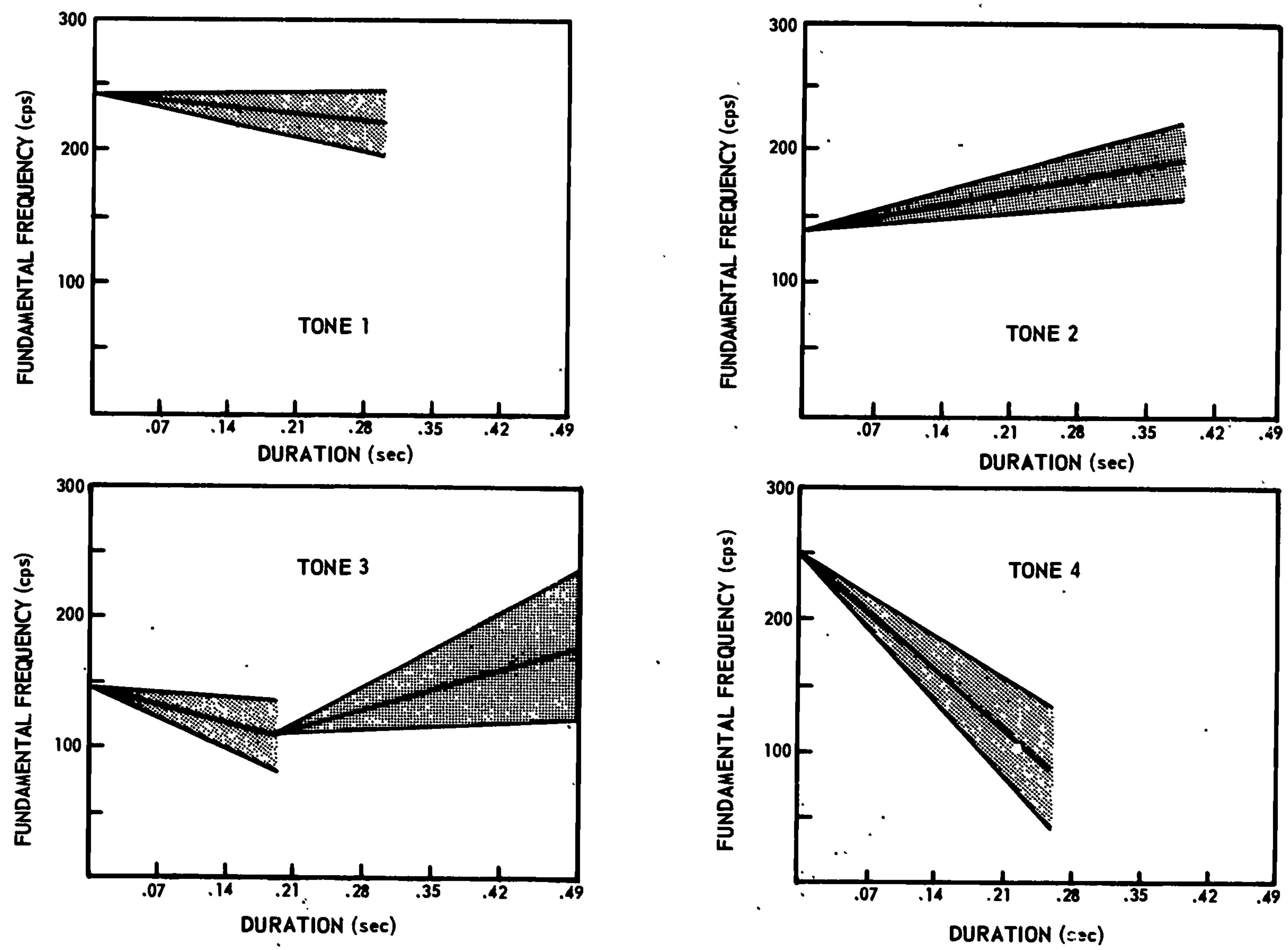


Figure 6. Emphatic Tones (±1-Sigma Limits)

informants in this study; that is, they set the statistically predicted endpoint frequencies one standard deviation on either side of our normalized contouroid.

Figure 6 shows that tone 1, tone 2, and the first part of tone 3 all have approximately the same magnitude of expected variation, whereas the latter segments of tones 3 and 4 show decidedly greater variability.

TONES AND THEIR NEUTRALS

Figure 7 shows the mean ±1-sigma slopes of the neutral tone, and it is obvious that the post-1 tone is the most unstable. Because of the ear's approximate 20-cps threshold, the tonal variation at the end of the post-1 neutral was cut off in the graph at 20 cps.

Figure 8 groups the four tones as they appear before a neutral. When contrasted with the emphatic tones (isolated) in Figure 4, it will be seen that in combination the tones are uniformly shorter (about 20 percent), and that while the second part of tone 3 is sharply curtailed in duration, it still has its characteristic dip. Figure 9 shows the four forms of the neutral tone itself and Figure 10, the effects exerted upon its contours by the particular tone preceding it. They are all, of course, significantly shorter than either the emphatic or combined full tone forms, and show, except for post-3, a sharp drop in pitch. The post-1 neutral starts at relatively the highest register and drops to the lowest of all the forms. The post-2 starting pitch is in register III, with the two remaining forms starting in register II.

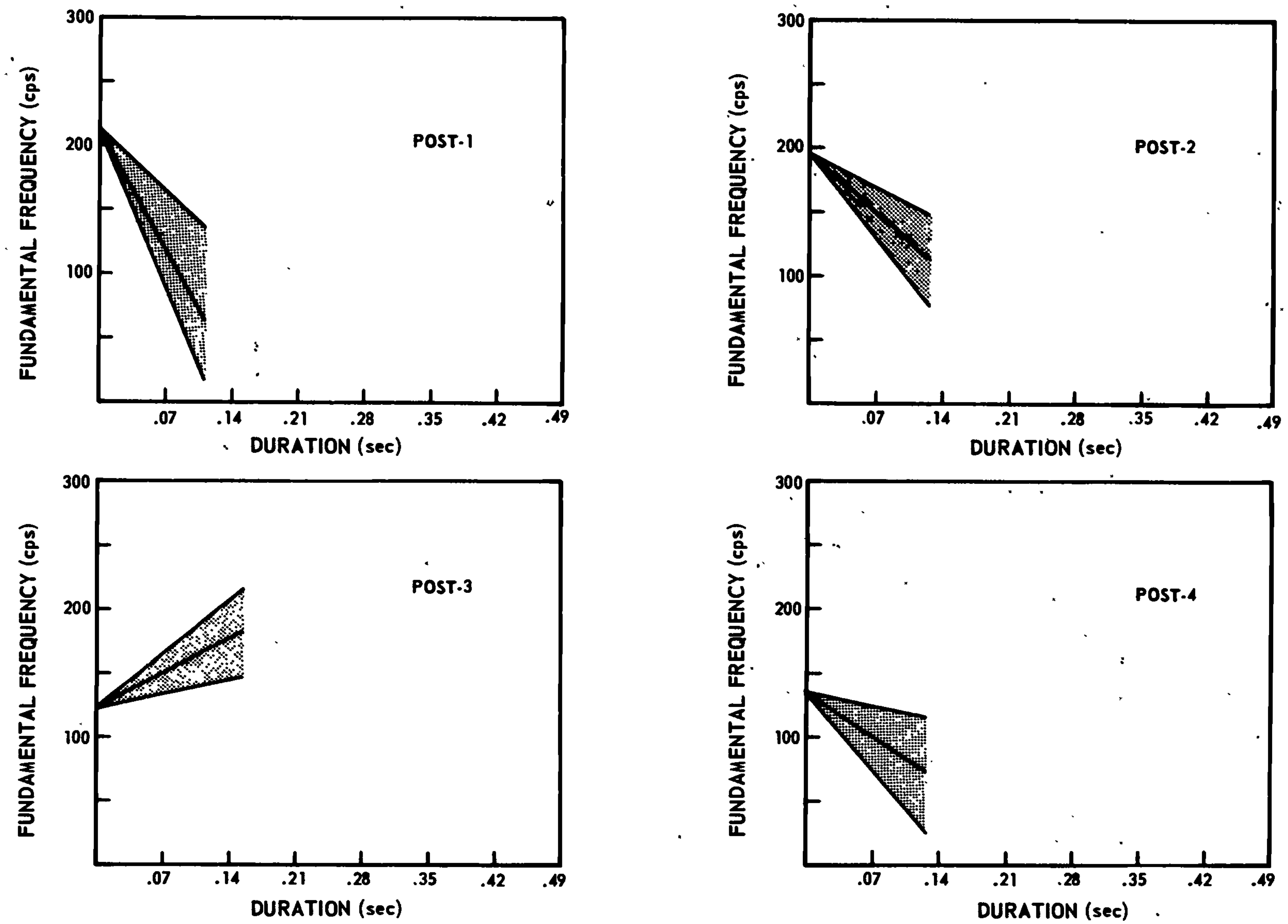


Figure 7. Neutral Tones (±1-Sigma Limits)

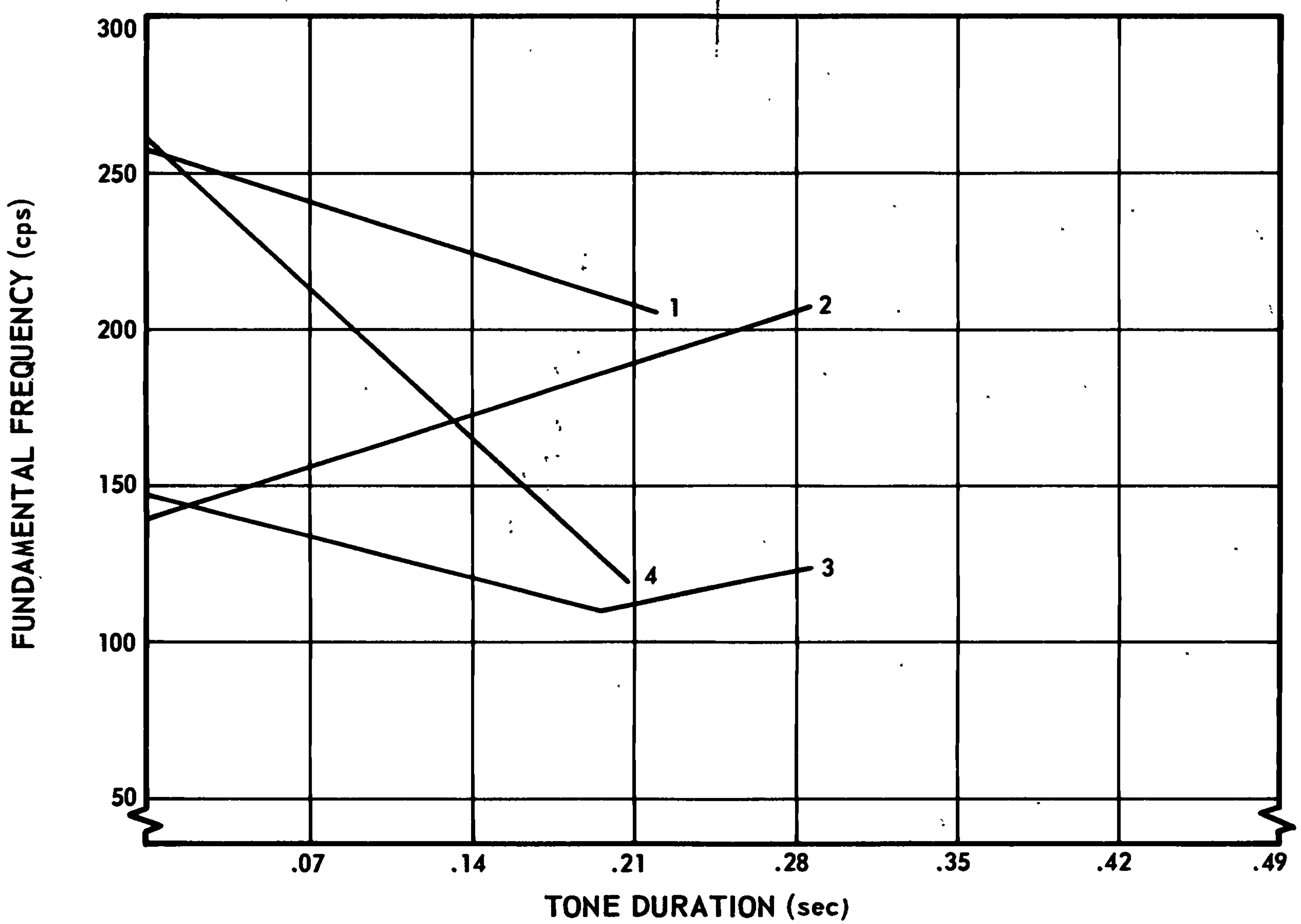


Figure 8. Contours of Tones in Preneutral Position

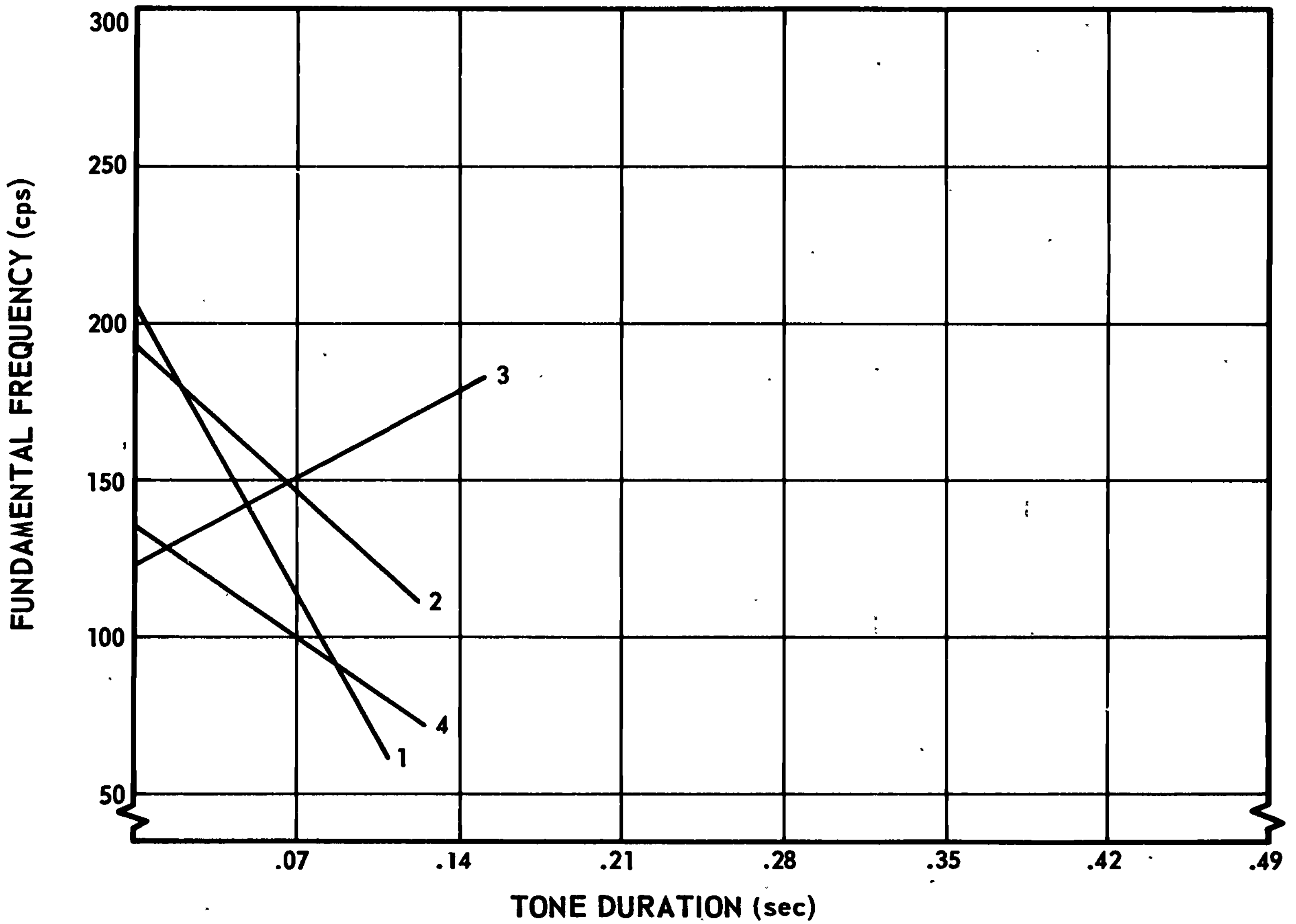


Figure 9. Neutral Tones

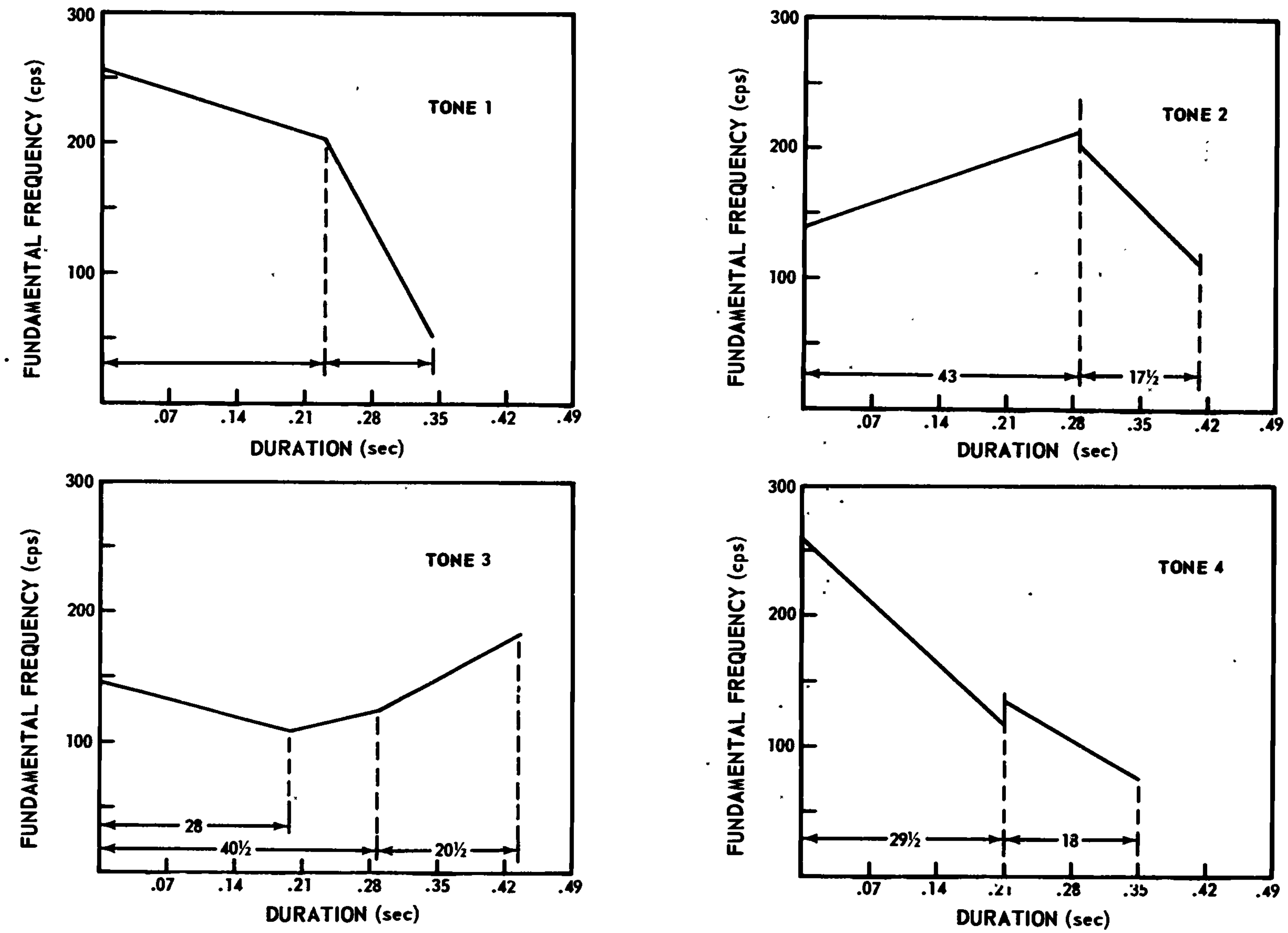


Figure 10. Tones Plus Neutrals

SANDHI EFFECTS ON TONEMES

Figures 11 through 20 individually consider the contours of the four tones as a function of the contour which precedes or follows. Perhaps some generalizations are in order regarding the whole set of displays:

1. Post-tones are longer than pre-tones.
2. Pre-tone forms show a greater spread of pitch regimes than post-tones.
3. Tones 1 and 4 appear to generate sharper slope families in the post- than in the pre-tone position.

Tone 3, of course, shows the most varied behavior, and the reader may make his own extensive comparisons from Figures 14 through 17. Where two types of contour are shown, as in Figures 15 and 16, the speaker group was about equally divided in their usage of the two forms.

By establishing pitch registers as we have done earlier, we may now classify the sandhi effects by superimposing a template upon the curves in Figures 10 through 19. These results are shown in Table 3.

From inspection of Table 3, it appears that tone 2 is the most stable form of all, retaining its 24 contour throughout.

Tone 4 shows a consistent pattern of a lower end register when following another tone than when preceding.

Both tones 1 and 4 start higher in isolation or when paired with a neutral than when spoken with any other tone.

Tone 3 shows greatest variability of form, and out of nine environments displays some nine different contours.

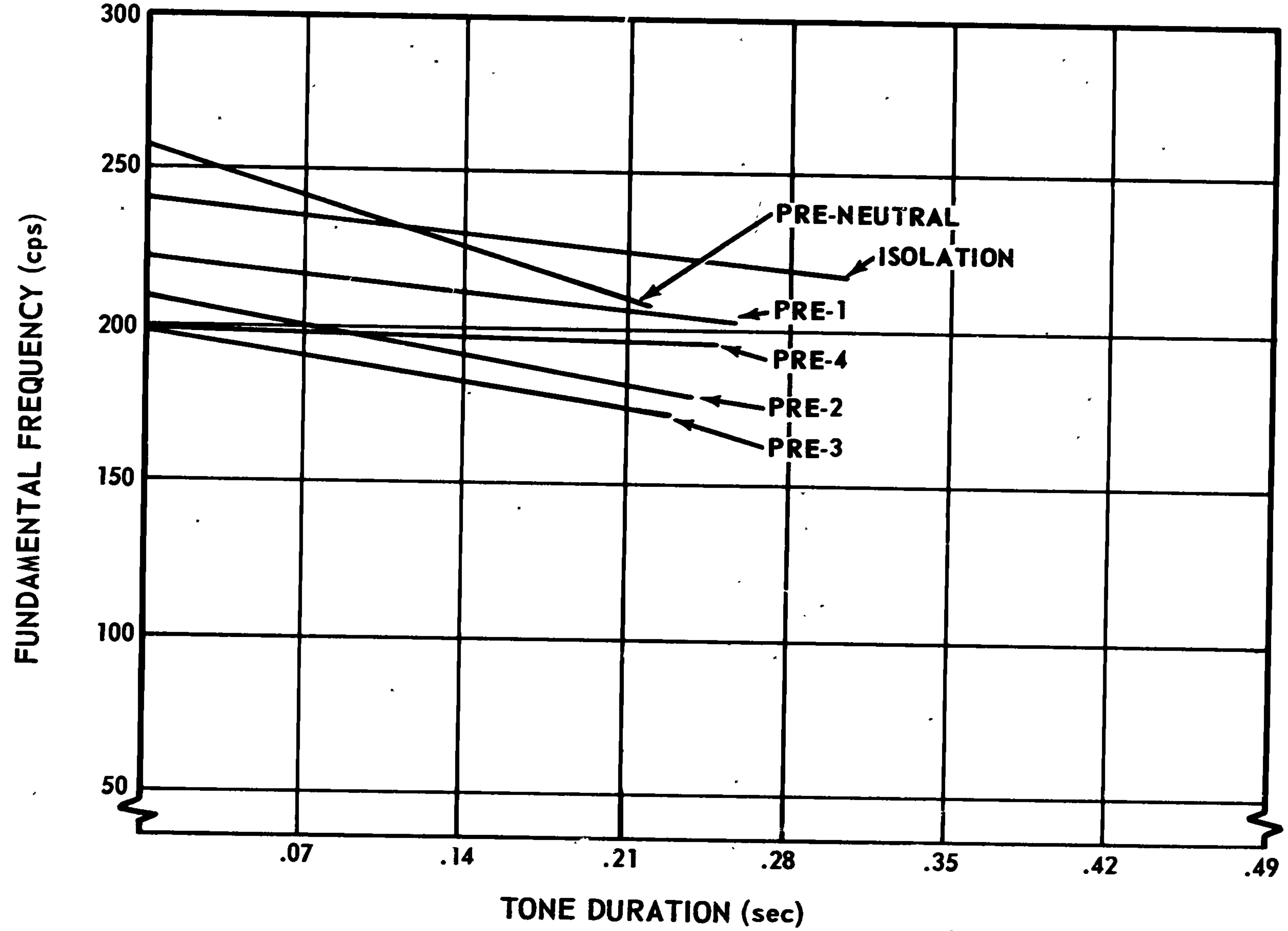


Figure 11. Tone 1 Contouroids -- Isolation, Pre-Neutral, 1, 2, 3, 4

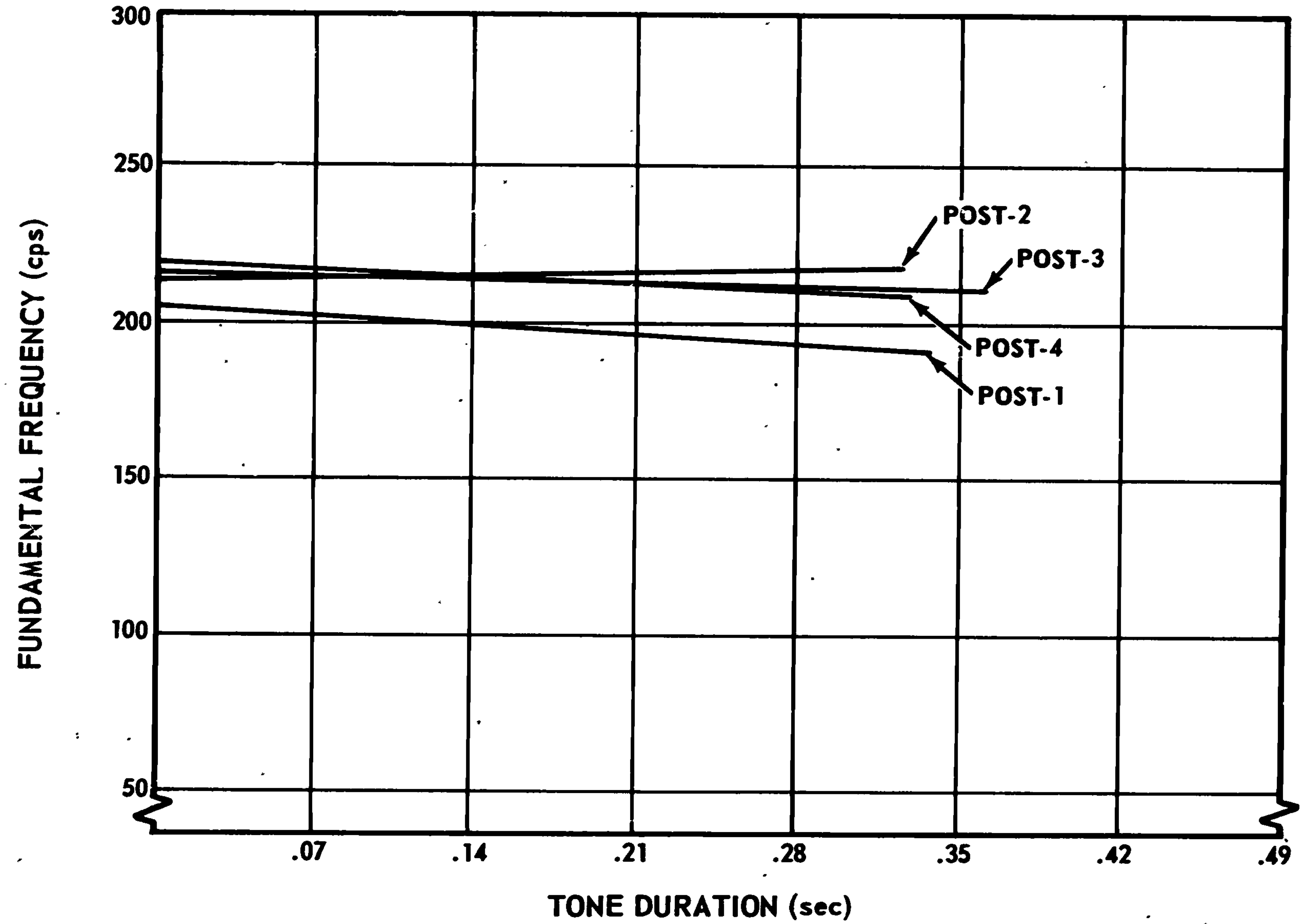


Figure 12. Tone 1 Contouroids -- Post 1, 2, 3, 4

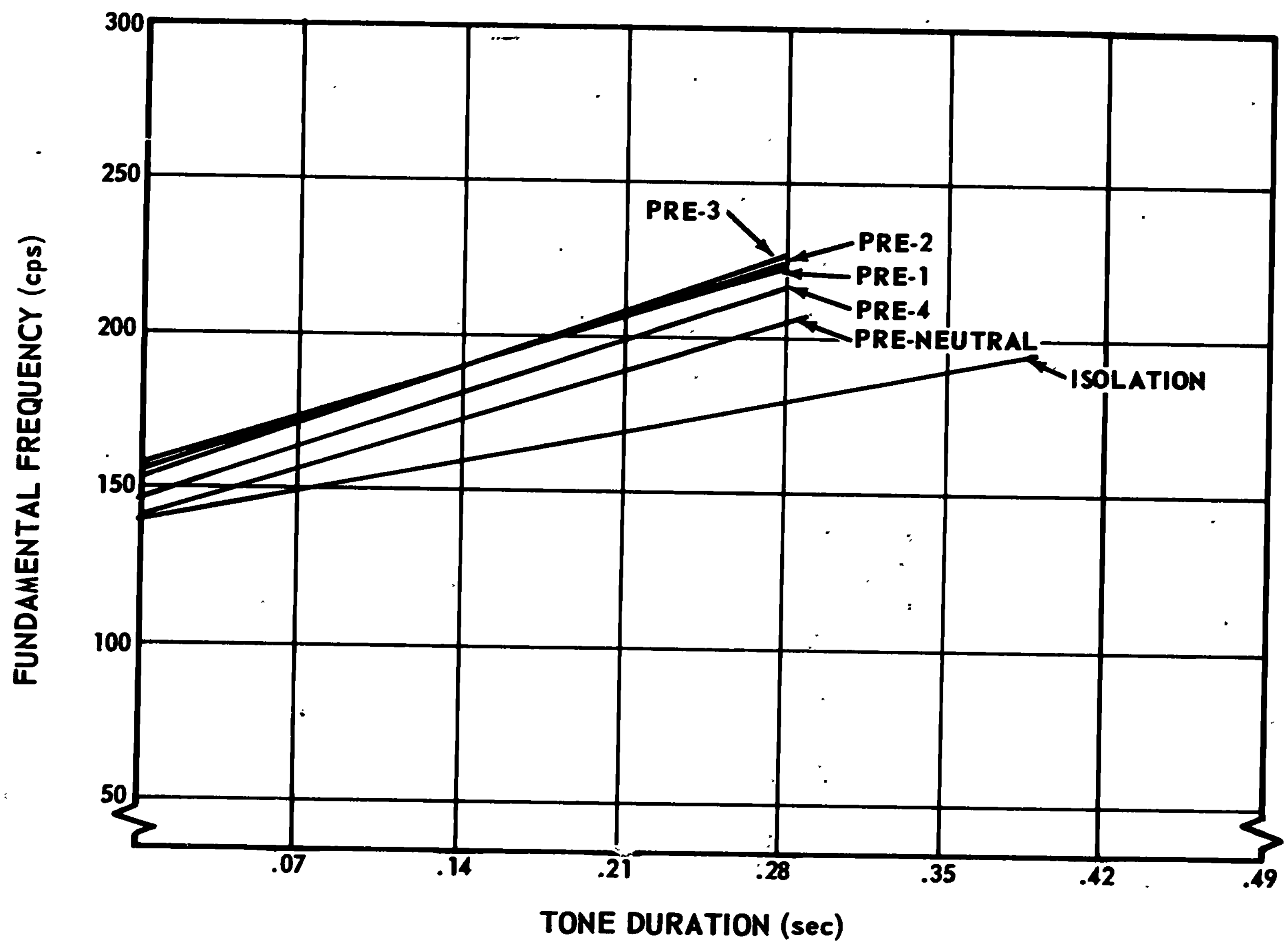


Figure 13. Tone 2 Contouroids -- Isolation, Pre-Neutral, 1, 2, 3, 4

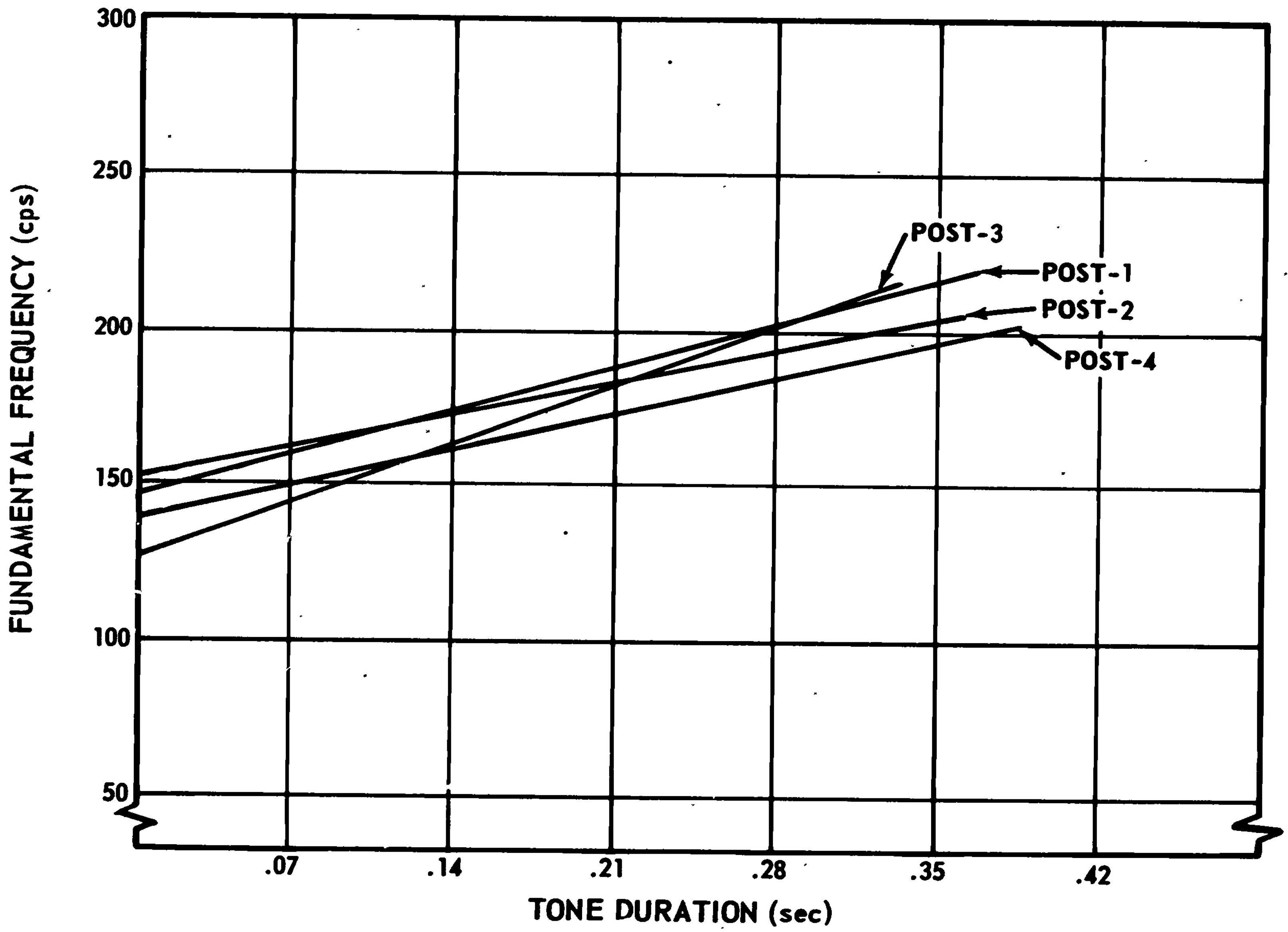


Figure 14. Tone 2 Contouroids -- Post 1, 2, 3, 4

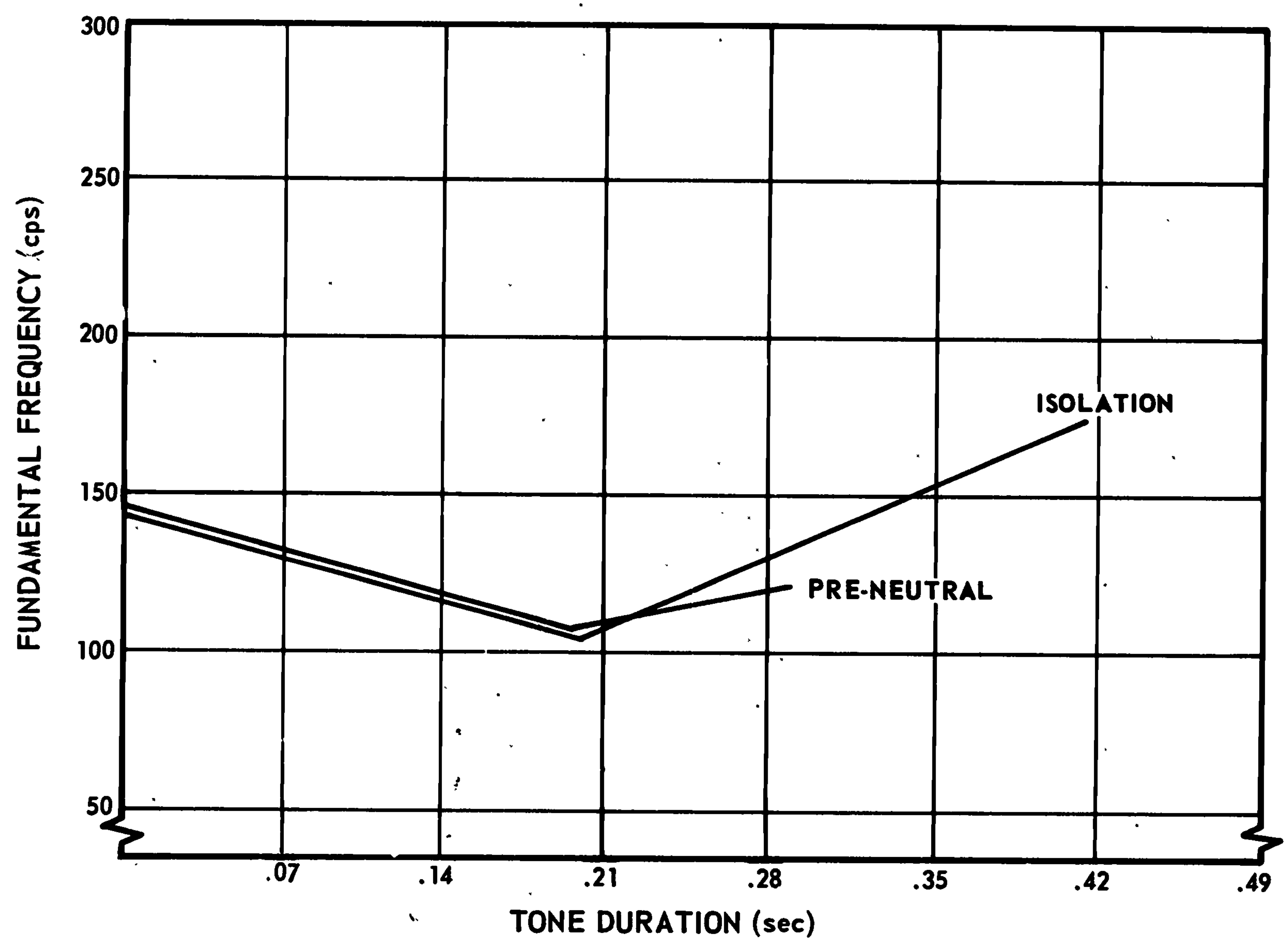


Figure 15. Tone 3 Contouroids -- Isolation, Pre-Neutral

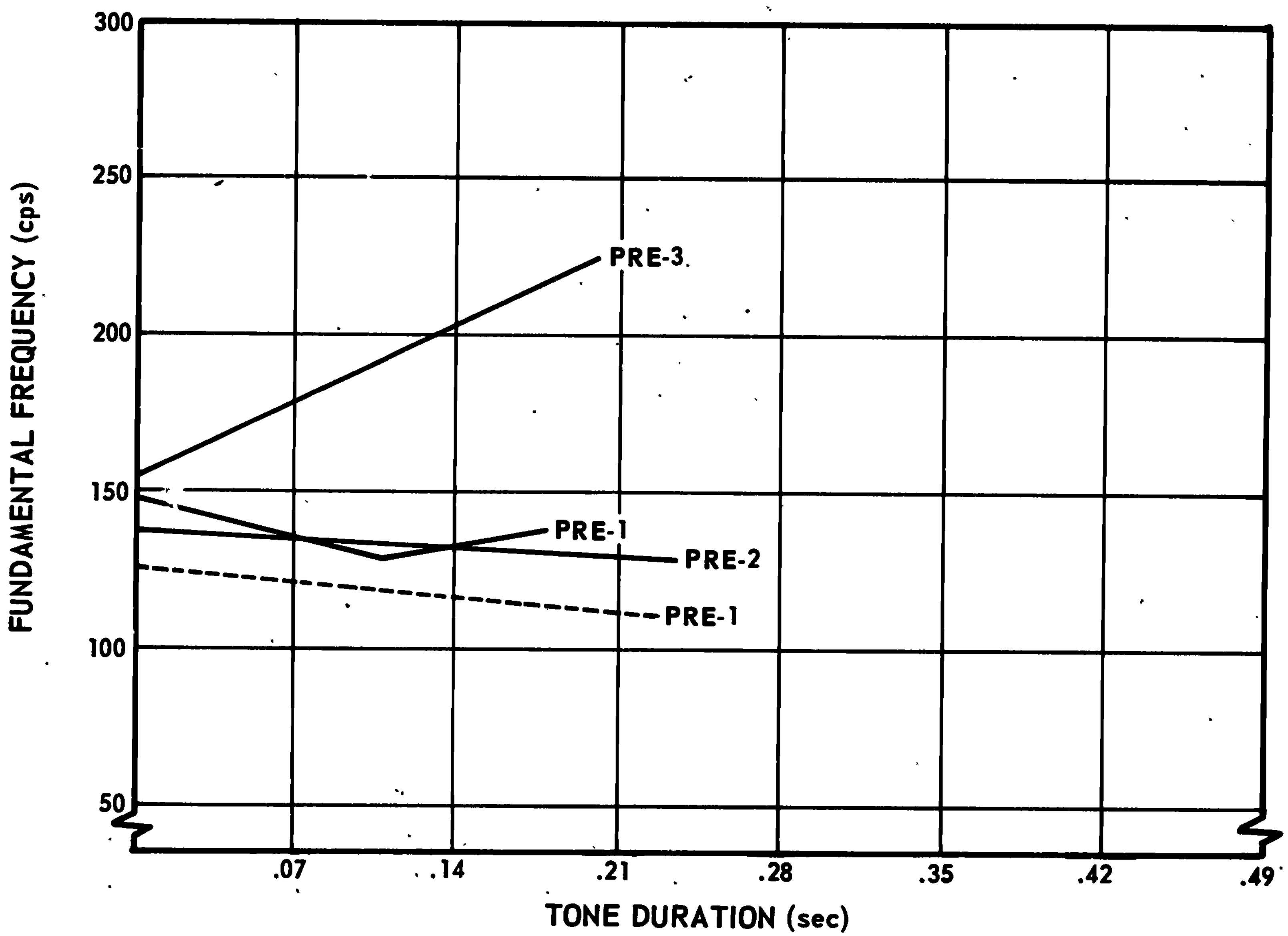


Figure 16. Tone 3 Contouroids -- Pre 1, 2, 3

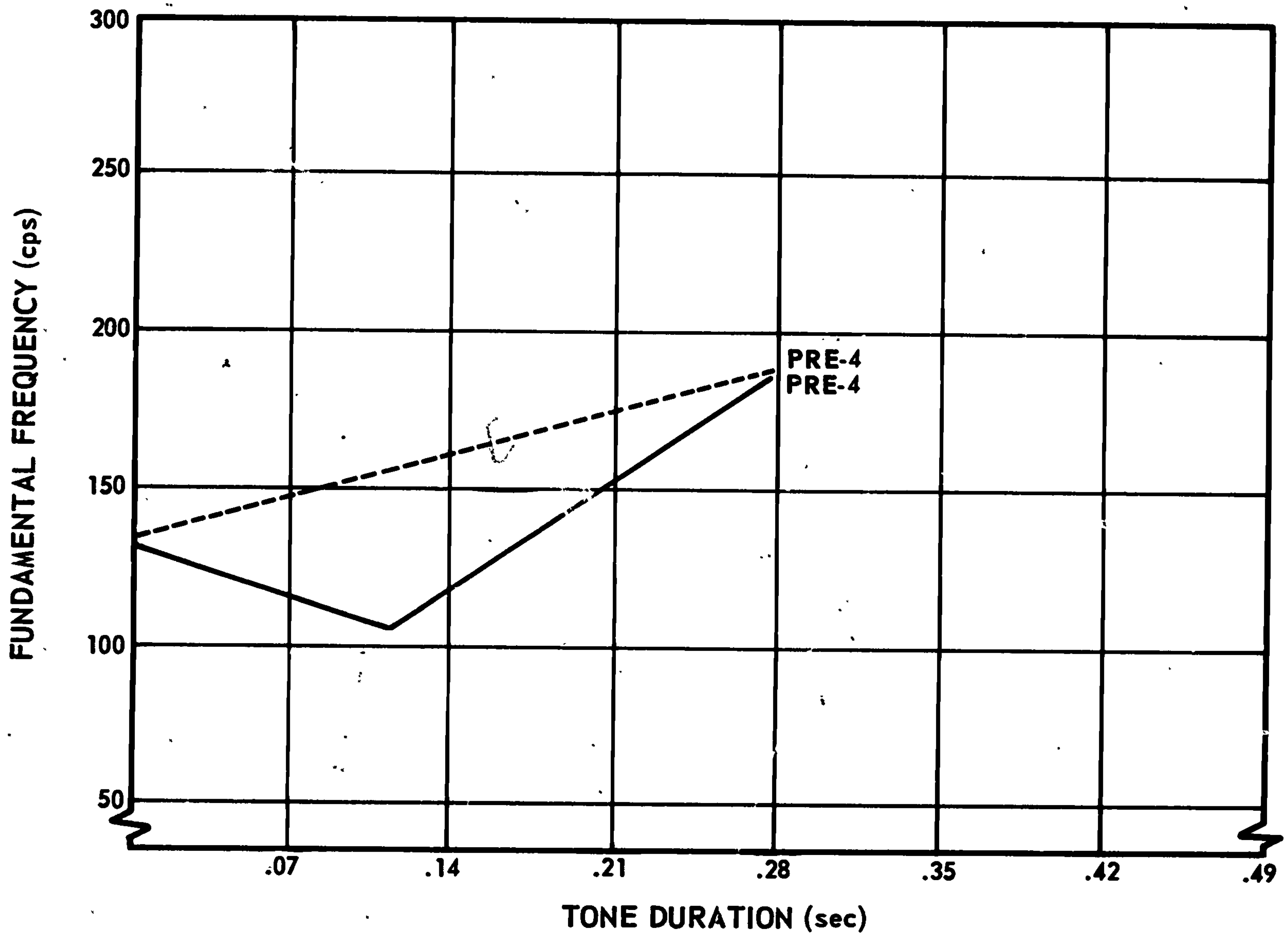


Figure 17. Tone 3 Contouroids -- Pre 4

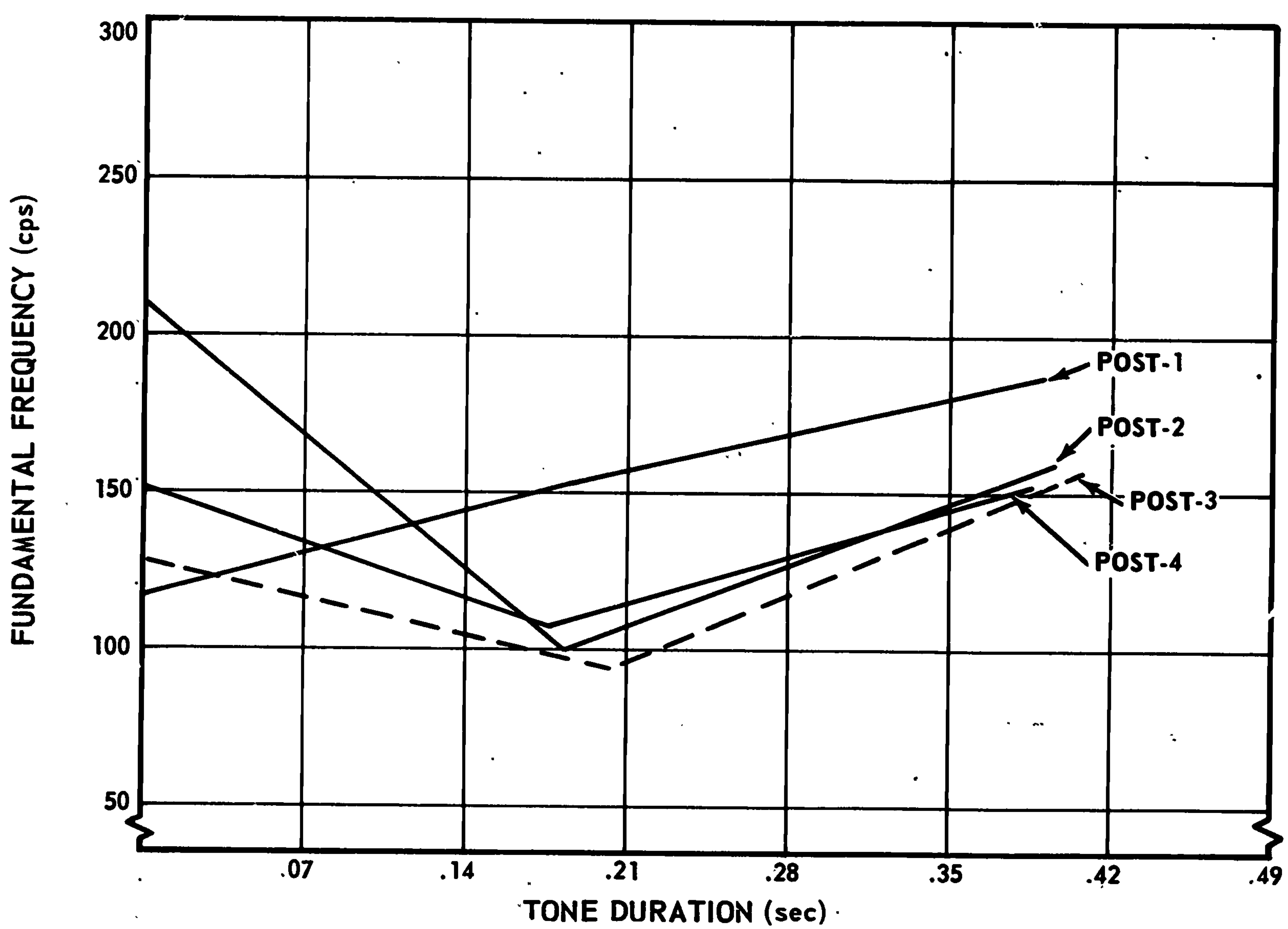


Figure 18. Tone 3 Contouroids -- Post 1, 2, 3, 4

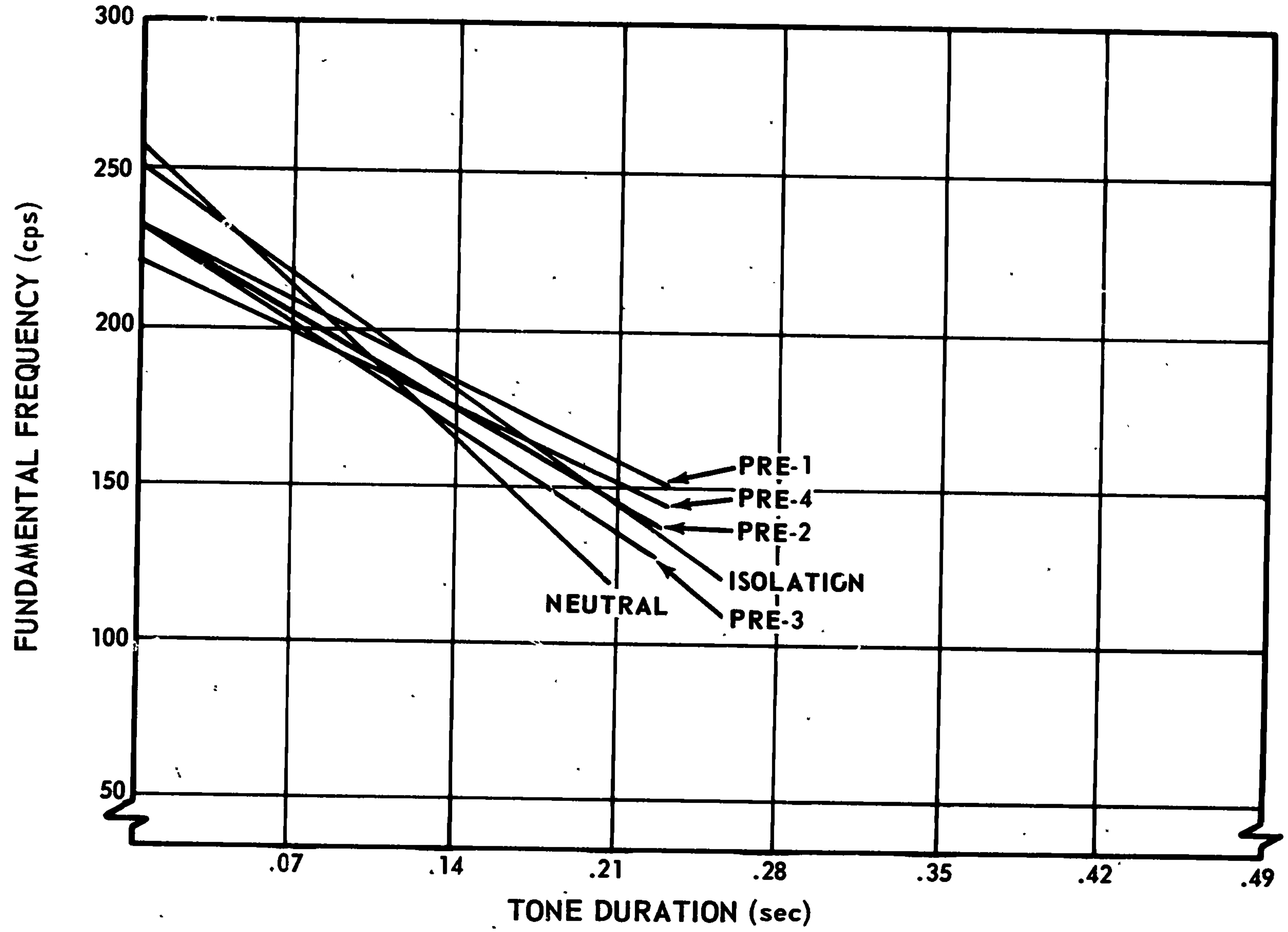


Figure 19. Tone 4 Contcuroids -- Isolation, Pre-Neutral, 1, 2, 3, 4

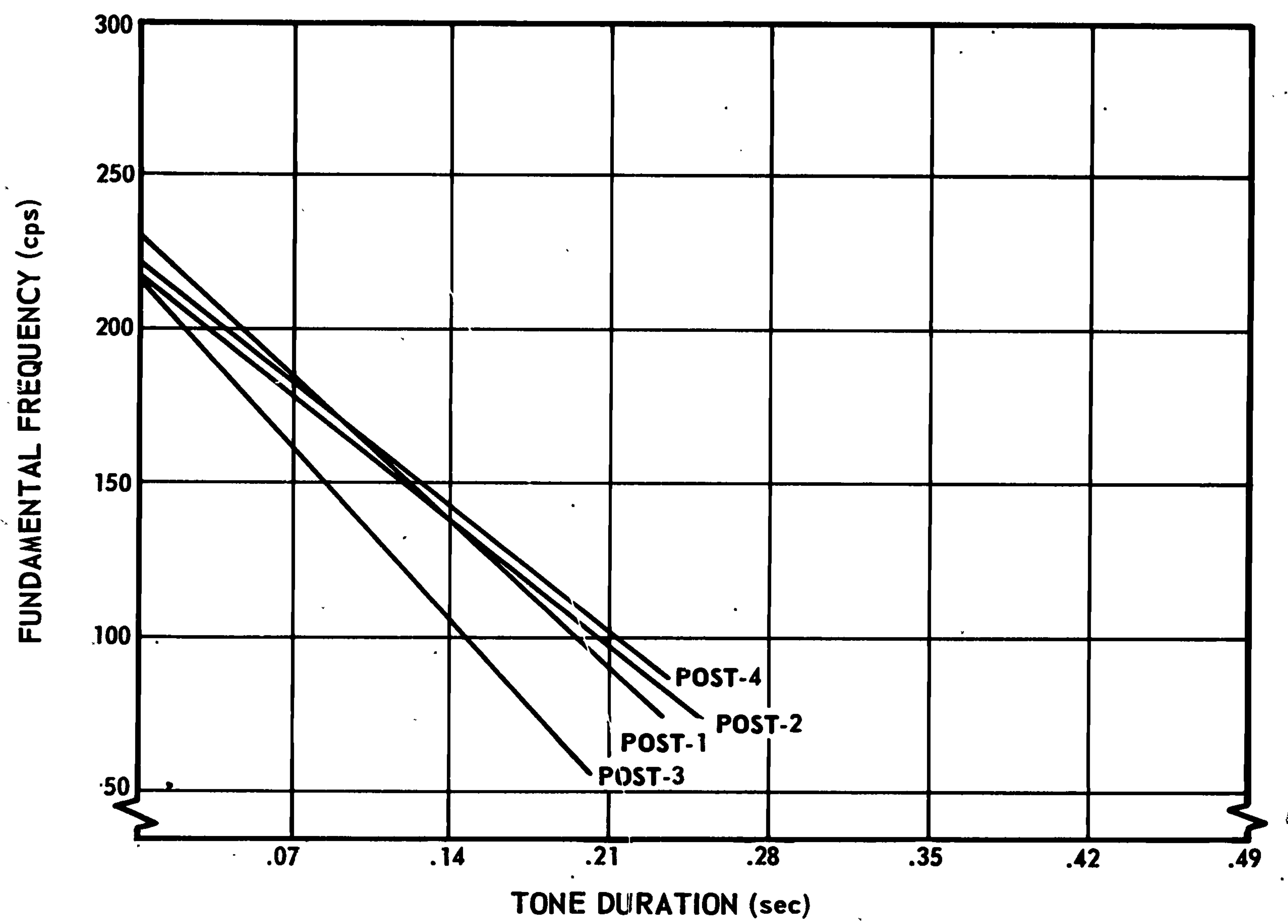


Figure 20. Tone 4 Contouroids -- Post 1, 2, 3, 4

Table 3

CONTOURS OF TONES IN COUPLET COMBINATION

Tone	Tones Exerting Effect							
	1		2		3		4	
	Before	After	Before	After	Before	After	Before	After
1	44	43	43	44	43	44	44	44
2	24	24	24	24	24	24	24	24
3	21 222*	13	22	213	24	313	23 213	313
4	42	41	42	41	42	41	42	41
Neutral		41		31		23		21

Tone	Pre-Neutral Contour
1	54
2	24
3	212
4	51

*Has a falling-rising contour, but excursion is held within the second register.

DURATION OF TONEMES

Durations of tonemes, determined from sonagrams, are shown in Table 4. Measurements were made from the start of the initial sound to the last detectable harmonic trace, disregarding occurrences of nonphonemic breath expulsion following pronunciation of the word.

Of interest to linguists at this juncture is a test of the hypothesis that tones used in combination actually change their durations, depending upon what they have been paired with. For instance, is tone 1 longer or shorter when paired with another tone 1 than when followed by tone 3?

An IBM QUIKTRAN computer program using double analysis of variance compared durations of each toneme as affected by its various environments to determine significant differences. These results are displayed in Tables 5 through 10.

In these tables L means significantly longer, S means shorter, and NSD means not significantly different in duration. The characterizing words refer to the tones shown in environments listed at the side of the table; thus the entry

	1 - 1
1 - 0	S

would be interpreted to mean, "Tone 1 in a pre-neutral setting is shorter than tone 1 in a pre-1 combination." A p point of 0.01 was selected as the necessary significance in evaluating F ratios. Entries refer to the

Table 4

DURATION OF MANDARIN
TONEMES IN ISOLATION AND IN COUPLETS

Position	Neutral	Tone (sec)			
		1	2	3	4
Emphatic		.302	.387	.412	.256
Preneutral		.221	.286	.288	.207
Pre - 1		.254	.281	.221	.232
- 2		.238	.267	.235	.226
- 3		.228	.274	.198	.228
- 4		.254	.283	.274	.232
Post - 1	.109	.335	.365	.390	.235
- 2	.130	.323	.361	.377	.250
- 3	.150	.361	.333	.394	.202
- 4	.125	.330	.383	.379	.236

Table 5

EFFECT OF COUPLET ENVIRONMENT ON DURATION OF FIRST TONES

Initial Position

	1-1	1-2	1-3	1-4
1-0	S	S	NSD	S
1-1		S	NSD	S
1-2			L	NSD
1-3				S

Final Position

	2-1	3-1	4-1
1-1	NSD	NSD	NSD
2-1		NSD	NSD
3-1			NSD

Table 6

EFFECT OF COUPLET ENVIRONMENT ON DURATION OF SECOND TONES

Initial Position

	2-1	2-2	2-3	2-4
2-0	NSD	NSD	NSD	NSD
2-1		NSD	NSD	NSD
2-2			NSD	NSD
2-3				NSD

Final Position

	2-2	3-2	4-2
1-2	NSD	L	NSD
2-2		L	S
3-2			S

Table 7

EFFECT OF COUPLET ENVIRONMENT ON DURATION OF THIRD TONES

Final Position

	2-3	3-3	4-3
1-3	NSD	NSD	NSD
2-3		NSD	NSD
3-3			NSD

Table 8

EFFECT OF COUPLET ENVIRONMENT ON DURATION OF FOURTH TONES

Initial Position

	4-1	4-2	4-3	4-4
4-0	NSD	NSD	NSD	NSD
4-1		NSD	NSD	NSD
4-2			NSD	NSD
4-3				NSD

Final Position

	2-4	3-4	4-4
1-4	NSD	L	NSD
2-4		L	L
3-4			S

Table 9

EFFECT OF COUPLET ENVIRONMENT ON DURATION OF NEUTRAL TONES

Final Position

	2-0	3-0	4-0
1-0	S	S	NSD
	2-0	S	NSD
		3-0	L

Table 10

DURATIONS* OF EMPHATIC TONES

	2	3	4
1	S	S	L
	2	S	L
		3	L

*Since each of the four emphatic forms is different from the remaining three, the durations (from least to greatest) are ordered 4, 1, 2, 3. It is interesting to note that the lexical forms in the dictionary (Fenn) are ordered by magnitude in the same sequence, that is, most fourth tones, fewest third tones.

relation of tones listed on left side to those listed along top of table.

From Table 5 we may infer that when tone 1 is the second member of a couplet, it shows no duration effect from its neighbor. However, when it is the first member, its duration depends upon what tone follows.

In contrast to tone 1, tone 2 (Table 6) exhibits reverse behavior, being insensible to environment when in a preceding position, sensitive when in a following.

Like that for tone 1, the contour for tone 3 (Table 7) shows excellent stability when the tone is in final position. Because of the inconsistent contour in initial position, this environment was not tested.

It will be seen from Table 8 that the same sort of duration pattern obtains with tone 4 as with tone 2; that is, stability of duration is associated with the initial but not the final position in a couplet.

Our experimental evidence allows us to conclude from the relationships in the preceding tables that environment does affect the duration of tones under certain circumstances, both as a function of what tone is uttered and where it is placed in the couplet. Undoubtedly a much more complicated picture will emerge as words are joined in longer utterances.

REMARKS

Appropriate conclusions from each step of the experimental study, contained in the body of the report, are not repeated here. In general, however, in sandhi form, the four Mandarin tones show more variability than had been suspected, although good speakers establish remarkable consistency in producing these variant forms. It is not implied that the variations should be applied to teaching techniques with humans. In teaching computers, however, these data may represent the start of necessary data compilation.

Section 4

NOTES

1. Liu Fu, <u>Ssu Sheng Shih Yen Lu</u>, (Experiments on Four Tones), Ch'un Yi Press, Shanghai, 1924.

2. Wang, Wm. S-Y. and Li, R. P. "Machine Recognition of Mandarin Monosyllables," Report No. 3, ONR Contract No. Nonr-495(27), June 1964.

3. Theoretically, the four lexical Mandarin tones have been considered in isolation:

   Tone 1: The voice starts in high register, maintains constant pitch.

   Tone 2: The voice starts in mid-register, sweeps upward in pitch.

   Tone 3: The voice starts mid- or low-register, dips slightly, and sweeps upward in pitch.

   Tone 4: The voice starts in high register, falls in pitch.

   To these is added a fifth tone, a so-called "neutral" or "zero." Relatively short in duration, it cannot occur in isolation, takes its pitch and contour character from its spoken context, and is denoted by a zero in the number notation system. Some common equivalent notation systems are noted below:

MANDARIN TONE NOTATION SYSTEMS

	System			
Tone	Wade-Giles	Chao	Yale	Chao
First	1	˥	(wōrd)	55
Second	2	˧˥	(wórd)	35
Third	3	˨˩˦	(wǒrd)	214
Fourth	4	˥˩	(wòrd)	51
Neutral			(word)	

4. The following words were used in this study:

1.	吃	(ch'ih)	2-1	民生	(mín-shēng)
2.	來	(lái)	2-2	民權	(mín-ch'üán)
3.	走	(tsǒu)	2-3	民主	(mín-chǔ)
4.	去	(ch'ù)	2-4	民治	(mín-chìh)
1-0	吃了	(ch'īh-le)	3-1	好天	(hǎo-t'iēn)
2-0	來了	(lái-le)	3-2	好人	(hǎo-jén)
3-0	走了	(tsǒu-le)	3-3	好冷	(hǎo-lěng)
4-0	去了	(ch'ù-le)	3-4	好熱	(hǎo-jè)
1-1	家兄	(chiā-hsiūng)	4-1	費心	(fèi-hsīn)
1-2	家嚴	(chiā-yén)	4-2	費錢	(fèi-ch'ién)
1-3	家母	(chiā-mǔ)	4-3	費水	(fèi-shuǐ)
1-4	家父	(chiā-fù)	4-4	費事	(fèi-shìh)

5. Each speaker was asked to speak in front of the microphone in his or her natural conversational tone and to speak with a slight pause between each word (a combination being considered a word).

6. Nominally, the 50-5000 cps expanded-scale spectrogram used in this investigation gives a 25-2500 cps display spread over 4 inches. The actual instrument used produced an exact 4-inch width tracing field and a frequency response of 30-2500 cps. Prorating the response/mm, we arrive at 24.31 cycles/mm as a display factor to use in calculating a fundamental. The pattern length for the instrument was 2.21 seconds per 12 inches, a resolution of 0.00725 sec/mm.

   For convenience, factors of 25 cps/mm and 0.007 sec/mm were adopted as sufficiently fine to be consistent with other experimental errors. Fundamental frequencies were therefore calculated from

$$f = \frac{\Phi d}{I_n}$$

   where

   f = fundamental frequency

   Φ = frequency resolution display factor

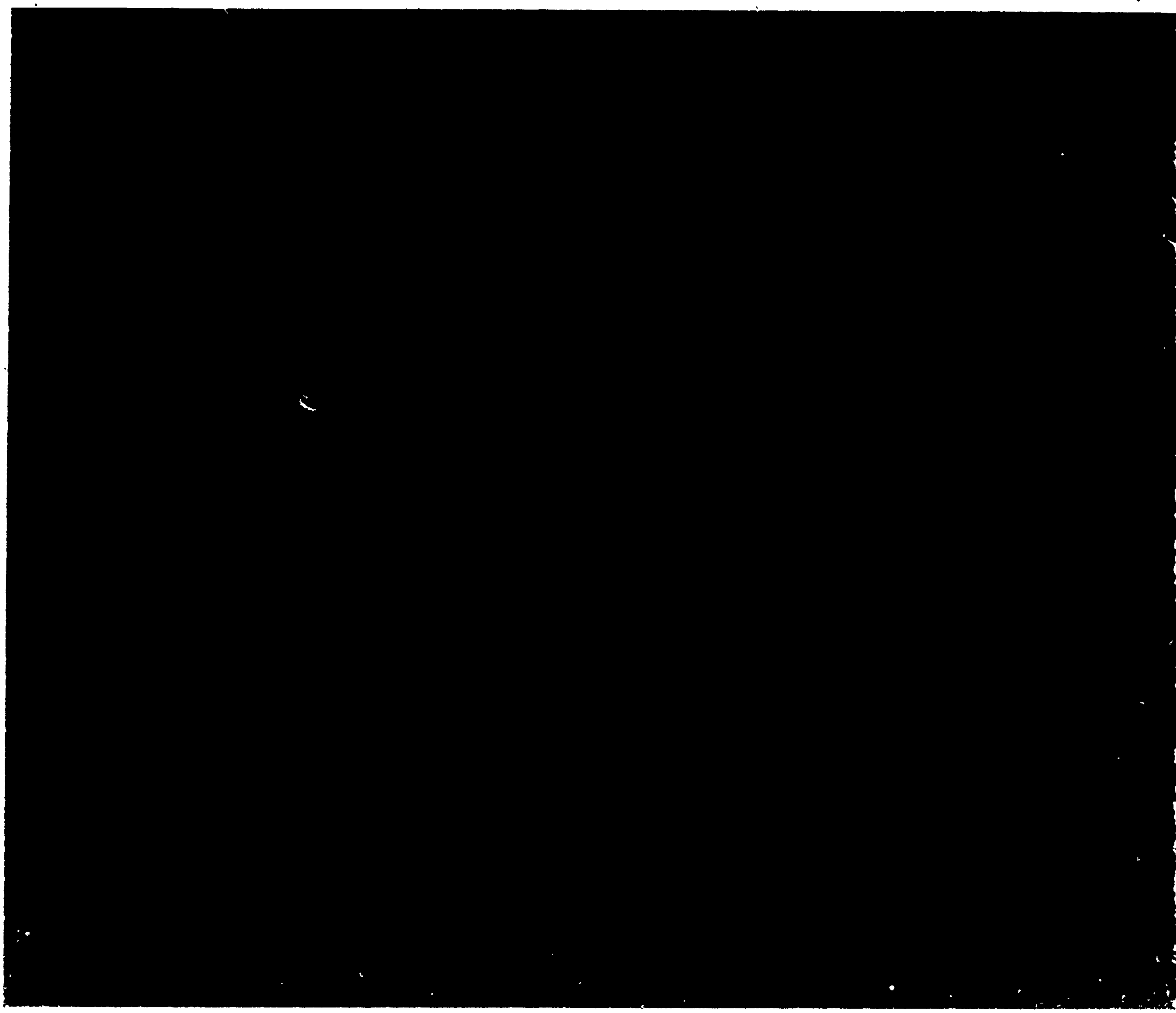
   d = distance in millimeters between midpoints of n harmonics

   I_n = number of intervening spaces between the n harmonics selected (always equal to n-1)

Connecting any two such calculated values by a straight line over a time t is the slope value characterizing the frequency change versus duration quantity of a tonally inflected speech unit.

7. Dreher, J. J., Intonation of Native and Acquired Languages, University of Michigan microfilm publication, 1951. Here is presented the "contouroid-vectoroid" concept, an application of machine analysis methods to tonemic contours. A vectoroid is defined as the average slope of the vocal fundamental between bounds of unvoicing or zero environment and a maximum or minimum of the fundamental trace. A contouroid is one or more contiguous vectoroids bounded by zero environment.

8. Pike, K. L., Tone Languages, University of Michigan Press, 1948. Chapter I of this work presents an exhaustive discussion of the mechanics of tone languages, including the question of the function of register in a contour system. See particularly the note on page 8 from Edward Sapir, regarding the structure of Lithuanian.



DOUGLAS

ADVANCED RESEARCH LABORATORY 5251 BOLSA AVENUE HUNTINGTON BEACH, CALIFORNIA