Aston, Carletta Hamilton

FUNDAMENTAL FREQUENCY AND DURATIONAL DIFFERENCES BETWEEN LEXICALLY IDENTICAL READ AND SPONTANEOUS UTTERANCES

City University of New York PH.D. 1982

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark _√_.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages _✓_
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) __________ lacking when material received, and not available from school or author.
12. Page(s) __________ seem to be missing in numbering only as text follows.
13. Two pages numbered __________. Text follows.
14. Curling and wrinkled pages _____
15. Other ______________________________________

FUNDAMENTAL FREQUENCY AND DURATIONAL DIFFERENCES BETWEEN LEXICALLY IDENTICAL READ AND SPONTANEOUS UTTERANCES

by

CARLETTA HAMILTON ASTON

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfilment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1982

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

May 14, 1982
date

Katherine S. Harris
Chairman of Examining Committee

May 14, 1982
date

[illegible]
Executive Officer

Lawrence J. Raphael

Katherine S. Harris

[illegible]

Supervisory Committee

The City University of New York

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS

Over the course of my graduate studies a number of people have contributed significantly to my development, both professional and personal.

First of all there is Katherine Harris, whose concern, encouragement and sensitivity as a strong role model and mentor have influenced me immeasurably.

Harry Levitt and Lawrence Raphael, with incisive comments and quiet support, have often helped clarify my murky thinking.

Osamu Fujimura and Lloyd Nakatani of Bell Laboratories introduced me to the world of computer manipulation of speech and in doing so opened up previously unknown exciting vistas.

My friends Claudette Feier and Esseline Reid have provided constant encouragement and moral support. Thanks to Dorothea Malcolm for the extra investment of time, patience and concern.

Last of all, but most importantly, there are the members of my family--Mother, Aunt Lou, Heather, Karriem and Julius--who were constantly supportive, even when they had no idea of what I was doing or why.

LIST OF TABLES

Table | Page

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

In order to produce intelligible, natural sounding synthetic speech, it is necessary to have parsimonious descriptions of all relevant linguistic and non-linguistic features. These linguistic and non-linguistic distinctions encompass segmental and suprasegmental variables. Features of the individual phonemes have been described in detail, so that highly intelligible synthetic speech can be realized. The ability to produce natural sounding synthetic speech rests on a complete description of the ways in which suprasegmental factors affect the speech signal. Primary among the suprasegmentals is the F_0 contour, which carries a large and multifaceted information load. Various approaches have been developed for describing neutral declarative F_0 contours which are appropriate for a limited set of utterances. Further work in this area will be focused on expanding F_0 generating rules to cover a wider range of syntactic structures, variation as a function of semantic emphasis and the effects of extralinguistic components.

One of the extra-linguistic factors which appears to markedly affect the speech signal is the level of formality of the speaking situation. Differences in durational, fundamental frequency and spectral characteristics have been noted as a function of different speaking modes. It is important to know how applicable are data obtained under one speaking mode to other speaking conditions.

The current investigation was designed to capture two distinct levels of formality by sampling speech that had been read and speech that had been produced spontaneously. An interest in subtle acoustic differences prompted the use of lexically identical utterances which contained none of the gross hesitation phenomena normally associated with spontaneous speech. In order to elicit these lexically identical utterances, speakers were taped during unplanned conversations. These dialogues were edited, transcribed and then read by the original speakers.

A perceptual task was devised to allow for elicitation of listeners' responses to the spontaneous/read (S/R) distinction under two conditions. First, as spoken, and second, with F_0 contours flattened to the average value for the utterances. Certain duration and F_0 parameters of the selected utterances were measured. These data were analyzed in an attempt to answer three specific questions:

1. Can listeners correctly identify an isolated utterance as having been read or spoken spontaneously?

2. Do listeners use the F_0 contour in making S/R judgments?

3. Are there consistent differences between read and spontaneous utterances in certain duration and fundamental frequency measures?

The long-range goal of this research is the development of a model encompassing all of the high and low level differences between the two speaking modes. Such a model should account for different degrees of spontaneity, as well as different levels of formality in reading. The test of the validity of such a model would be whether an algorithm could be derived from it for altering the perceived spontaneity or "readness" of utterances. This investigation was the initial step towards this long range goal.

CHAPTER II

REVIEW OF RELEVANT LITERATURE

The literature reviewed in this chapter is focused on three areas. The first and second areas are F_0 and duration. These are the two acoustic parameters which were examined in the current investigation of the S/R distinction. The third area is concerned with studies related to the S/R distinction. Included are investigations which have established differences between these speaking modes, as well as some work that points to potential differences between them.

Fundamental Frequency

Physiological Correlates of Fundamental Frequency

Phonation is generally accepted to occur as a result of certain myoelastic and aerodynamic forces working in concert (van den Berg, 1958). The myoelastic forces are associated with contraction of the intrinsic laryngeal musculature primarily and with activity of certain extrinsic laryngeal muscles secondarily. The activity of these muscles causes the vocal folds to be adducted and abducted, stretched and tensed. The aerodynamic forces are associated with the breath stream escaping from the lungs which forces the adducted folds apart and also contributes to their closing.

Fundamental frequency change is most closely related to contraction of the cricothyroid (CT) muscles. Increase in F_0 as a direct result of CT contraction is well documented

(Faaborg-Andersen, 1957; Katsuki, 1950; Lieberman, Sawashima, Harris and Gay, 1970). CT activity causes F_0 raising through an increase of longitudinal tension on the vocal folds. A less important contribution to raising of F_0 may occur through contraction of the suprahyoids which results in an increase in vertical tension (Borden and Harris, 1980).

Relaxation of the contracted CT seems to be the most important factor in lowering of F_0 (Sawashima, Kakita and Hiki, 1973; Collier, 1975; Erickson and Atkinson, 1976). There also appears to be an active, if minor, F_0 lowering component as a function of contraction of the strap muscles which alter the position of the larynx in the neck (Lieberman et al., 1970; Ohala, 1970; Atkinson, 1973; Sawashima et al., 1973; Baer, 1975; Collier, 1975; Maeda, 1976). The degree to which this active F_0 lowering mechanism is functional may be idiosyncratic and vary throughout a given speaker's F_0 range (Atkinson, 1978). If all muscular forces are held constant, changes in F_0 will occur as a function of subglottal pressure variation. Although Lieberman (1967) attributed F_0 declination primarily to a fall in subglottal pressure, the amount of this contribution was probably overestimated.

Linguistic Correlates of Fundamental Frequency

The physiological mechanisms that underlie changes in fundamental frequency of the voice are controlled by the linguistic requirements of the speaker. On the segmental level there are a number of factors which affect fundamental frequency. Each vowel is purported to have an intrinsic fundamental frequency which is associated with relative vowel height (Lehiste and Peterson, 1961). Although this acoustic characteristic of vowels is well-documented in production of isolated words and carrier phrase-type utterances, it may not be realized in connected narrative type discourse (Umeda, 1981).

In addition to intrinsic F_0s, the level of the F_0 peak associated with the vowel, the direction of movement of F_0, rate of movement and duration of movement are conditioned by the preceding consonant (House and Fairbanks, 1953; Lehiste and Peterson, 1961; Maeda, 1976; O'Shaughnessy, 1976). These differences, which can be explained on the basis of events occurring at the level of the larynx have been found to serve as cues in the voice/voiceless distinction (Haggard, Ambler and Callow, 1970).

F_0, along with duration and intensity, has been shown to vary with degree of lexical stress (Fry, 1958; Lieberman, 1960; Morton and Jassem, 1965). F_0 and duration appear to have primary cue value in perception of lexical stress, although their relative contribution is dependent upon a number of variables. These variables include position in the utterance and idiosyncratic speaker and listener differences (Nakatani and Aston, 1981).

It is logical to assume that F_0 would be an effective cue to sentential stress, although the notion of sentential stress is less clear cut than that of lexical stress. Bolinger (1958) has suggested that the concept of stress, at the sentence level, associated with force of utterance and intensity, should be replaced by the notion of pitch prominence.

Fundamental Frequency Contour

It is recognized that intonation is a perceptual description of the instrumentally defined F_0 contour. However, this distinction is not always maintained in the literature. Therefore, in the following review the terms F_0 and intonation will be used interchangeably.

Any characterization of the F_0 contour is complicated by a number of factors. First there are the various correlates of F_0 which are responsible for considerable variation--intrinsic vowel F_0, phonetic environment and syntactic structure. Second, the F_0 contour appears to have a number of independent functions. Third, as a consequence of these first two factors, there is no generally accepted system for describing F_0 contours.

Functions of the F_0 Contour

Daneš (1960) divided up intonation function into two primary and two secondary categories. The two primary structural functions are the delimiting or integrating of words within a sentence (breaking it up into phrases) and the signalling of theme and rheme (or old versus new information). These two operations are syntactic and semantic respectively. According to Daneš' breakdown, intonation also has two secondary or modal functions. The first of these is to indicate whether the sentence is a statement or a question; the second modal function being the indication of the affective state of the speaker.

Each of these four functions is embodied in a single F_0 contour and one of the fundamental tasks of the investigator of intonation is to isolate those aspects of the contour associated with each. By necessity research proceeds as though it is possible to consider them individually. Different investigators place varying amounts of emphasis on the different functions of intonation.

In addition to the four functions described by Daneš (1960) the F_0 contour serves a phonemic function when movement of the F_0 contour allows for distinction between voiced and voiceless sounds. It also gives lexical information through the role that it plays in signalling stress differences. A complete listing of all of the functions of the F_0 contour would also have to include providing information as to the speaker's sex, age (Collier, 1975) and linguistic background (Lieberman, 1967).

Wang (1972) points out that the intonation contour is produced within the constraints of certain operating characteristics of the speech production mechanism which are highly idiosyncratic. This consideration, in combination with all of the functions attributed to intonation, make the F_0 contour a highly fruitful, if complicated area for investigation. Acoustic effects associated with these various functions will be described in a later section on "Acoustic Studies of the F_0 Contour."

Impressionistic Descriptions of Intonation

Level analyses. Much of the earlier work on intonation grew out of pedagological needs of teachers of English as a second language. Although a large number of systems emerged, they can be grouped into two general categories. The first large category that covers the work of American linguists is based on an attempt to describe intonation as a series of phonemic pitch levels. The early approaches of Sapir (1921), Bloomfield (1933) and Wells (1945) were attempts to describe the suprasegmental aspects of pitch and the resulting intonation contour in a form analagous to segmental analysis. Intonation was considered a morphemic entity comprised of a series of pitch phonemes.

These early works were followed by the more detailed "level analyses" of Pike (1945) and Trager and Smith (1951). Pike attempted a complete description of one dialect of American English using four pitch phonemes or levels. Trager and Smith included various types of juncture in their analytical system which treated pitch and stress as independent.

through the assignment of stress placement rules, Chomsky and Halle (1968) and Halle and Keyser (1971) have attempted to predict sentence stress patterns from surface structure. Bresnan (1971) expanded the Chomsky and Halle system to include deep structure considerations. These three stress assignment systems are assumed to be part of linguistic competency only, and not necessarily reflected in physical reality.

Liberman (1975) developed a metrical system for predicting primary stress of neutral intonation contours. Although originally no claims were made as to the shape of the F_0 contour, a subsequent computer program was devised for generating F_0 contours within the framework of this theoretical model (Pierrehumbert, 1981).

Bolinger (1958, 1972) takes issue with these last three approaches on two counts. First of all, he feels that a phonemic approach to intonation analysis does not make sense because of the semantic discontinuity of phonemes as well as their relationship to a number of different articulatory correlates. Second, he believes that syntactic considerations only indicate which syllables are likely to be stressed. It is semantic and emotional highlighting which determines final stress patterns.

<u>Contour analysis</u>. The so-called British school of intonation uses contours or a series of configurations. Palmer (1922) devised a system which catalogued configura-

tions for three parts of the intonation contour--the nucleus, the head and the tail.

A somewhat more simplistic approach was that of Armstrong and Ward (1926). They described two possible intonation contours--Tune I and Tune II--which were assigned to an utterance depending on whether it was a statement, command, "yes-no" question or interrogative-word question. Other systems have been devised by Jassem (1972), Halliday (1967) and Crystal (1969). These contour analysis systems have suffered from having either too little or too much detail.

O'Connor and Arnold (1963) provided a thorough description of colloquial English intonation using a graphic system of small and large dots. They described a variety of tunes associated with the head, nucleus and tail of intonation contours of different sentence types.

<u>Acoustic/Perceptual Studies of Fundamental Frequency Contours</u>

<u>Neutral intonation contours</u>. There are numerous references in the literature to contrastive stress and emphatic stress which are functional at the sentence level (Bolinger, 1958; Lehiste, 1970; Berman and Szamos, 1972; Hirst, 1974). Implicit in the notion of these different types of stress is that a baseline exists against which a given pattern can be judged. Such a basic pattern is referred to as a "normal" or "neutral" intonation contour (O'Shaughnessy, 1976; Stockwell, 1971). O'Shaughnessy suggests that "...the speaker apparently adopts a strategy

of deviating from the 'normal' pattern to signal some linguistic or phonetic message to the listener." (p. 397).

The concept of main sentential stress is central to characterization of a neutral contour. Although this is a particularly murky concept, numerous references to main stress are made in the literature. Bolinger (1958) and Chomsky (1971) both suggest that main stress tends to occur toward the end of the utterance, although the mechanisms which they propose for its assignment are vastly different. Bolinger's mechanism, related to "new" versus "old" information is semantic in nature. Chomsky, on the other hand, ascribes stress assignment to "...special grammatical processes of a poorly understood sort." (p. 199). Atkinson (1973) found that listeners tend to assign main stress toward the end of an utterance even in the absence of typical F_0 behavior associated with main stress. Uldall (1960) has suggested that neutral intonation contours have relatively narrow ranges and less change in F_0 toward the end of the utterance.

At least one linguist rejects the notion of the neutral intonation contour (Schmerling, 1974) on the basis that neutralness could only be assigned to a sentence or utterance with no meaning. It is unclear as to whether this neutral pattern is supposed to be manifest acoustically or whether it is simply a mental image of a normal utterance (Coker and Umeda, 1971).

Declination. It is possible to look at localized attributes of the F_0 contour or to emphasize more global features. The best documented global feature of F_0 contours is declination. The declination effect was described in 1973 by t'Hart and Cohen as a gradual downdrift of pitch from the beginning to the end of an utterance. Possibly as a consequence of the fact that it occurs in a large number of unrelated languages (Pierrehumbert, 1978), declination has been considered a language universal which is performance based (Lieberman, 1967; Maeda, 1976). Through a series of experiments using reiterant speech techniques (Liberman, 1977) it has been demonstrated that this downdrift has a place in the linguistic system of American English (Breckenridge, 1978).

Earlier work on declination characterized this phenomenon by use of a negatively sloping line drawn by eye through or connecting the valleys of the intonation contour (Maeda, 1976). Later work has suggested that in addition to the "baseline" declination, a line drawn connecting the peaks of the F_0 contour is also negatively sloping. If both the topline and the baseline are drawn for a single utterance it will be seen that the topline has a more negative slope (O'Shaughnessy, 1976; Cooper and Sorenson, 1981). The space described between these two converging lines can be considered to be the intonation envelope (Cooper and Sorenson, 1981).

Because of proven perceptual relevance of the topline (Breckenridge, 1978) and the relative ease with which the topline can be drawn as compared with the baseline, recent efforts have attempted to describe the topline. Pierrehumbert (1981) describes a computer program for synthesis of F_0 contours. This program is based on the assumption that intonation contours can best be described by target values (pitch peaks) and transition rules that describe the movement of F_0 between the peaks.

Cooper and Sorenson (1981) have attempted to characterize the topline for a variety of sentences with different syntactic structures. Through comparison of F_0 data with several possible models, they concluded that F_0 topline could best be described by a single line connecting the second through the final peaks, with the initial peak located somewhere above the line. This system was seen to be more or less effective for sentences containing complement clauses, as well as simple sentences. It worked less well when adjectives and adverbs were included, in the presence of subordinate restrictive relative clauses and parenthetical clauses. These other syntactic permutations created elbows in the topline either going upward or downward depending on the utterance.

All of the topline contours displayed by Cooper and Sorenson were negatively sloped or falling. There are a limited number of references in the literature to rising F_0 contours (Palmer, 1922; Crytal, 1969; Pierrehumbert, 1980). They can be considered atypical since the vast majority of utterances have falling toplines.

Localized attributes. In addition to the global aspect of declination much emphasis has been placed on describing the more localized attributes which cover one or two words at most. A series of experiments on Dutch intonation have isolated a number of distinctive patterns that are superimposed on the gradually declining baseline (Cohen and t'Hart, 1967; Collier and t'Hart, 1972; t'Hart and Cohen, 1973). These patterns include the "hat pattern", the "valley pattern" and the "cap pattern." These patterns each reflect different F_0 rise, fall and peak behavior.

Maeda (1976) investigated F_0 contours of American English and found that he was able to characterize a limited number of declarative sentences using the five attributes of baseline, rise, lowering, peak and a rise on F_0 plateau. Sentences synthesized with intonation contours assigned on the basis of these attributes were perceptually acceptable.

Much of the work on localized attributes of the intonation contour has been concerned with various short duration movements of the F_0 contour and their relationship to various functions associated with it. The major accepted role of intonation is in syntactic grouping (Lea, 1972; Olive, 1975). Although the syntactic structure of an utterance can certainly be derived from monotone speech, the intonation contour can aid in the process of parsing into constituent structure (Gleason, 1961; Lea, 1973; Wingfield and Klein, 1971). This is particularly true when ambiguity is present (Streeter, 1978).

The marking of phrase boundaries by F_0 is usually accomplished by a sharp rise at the beginning of the phrase and a sharp fall at its close (Lea, 1973; O'Shaughnessy, 1976). The terminal fall at the end of a sentence is usually lower in frequency than the previous falls (O'Shaughnessy, 1976; Maeda, 1976). Continuation rises often mark nonterminal falls within sentences, signalling that more is to come (Daneš, 1960; Coker and Umeda, 1971; O'Shaughnessy, 1976).

Related to this syntactic segmentation is the notion that F_0 may be functional in indicating juncture for phonological units larger than the sentence (Lehiste, 1979; Umeda, 1978).

Semantic highlighting is another well-recognized function of the F_0 contour. This highlighting occurs with less predictable words or anaphoric referents (Daneš, 1960; Halliday, 1967; Atkinson, 1973; O'Shaughnessy, 1976). Presumably the more important, or unusual a word is, the higher the F_0 (O'Shaughnessy, 1976). It has been suggested that syntactic information is relayed by the direction of F_0 movement, while semantic information is reflected in the degree of movement (Adams, 1969; O'Shaughnessy, 1976).

Localized attributes are also affected by the phonetic influences already noted. O'Shaughnessy (1976) found that number and closeness of F_0 accents can affect the height of the F_0 peaks. When a number of accents are close together, and when there is a relatively large number of them, peaks will be lower.

In addition to the linguistic determinants mentioned above, F_0 also varies with certain non-linguistic factors. Listeners have been shown to be able to identify speakers (Abberton and Fourcin, 1978), and emotional modes (Uldall, 1960; Lieberman and Michaels, 1962), on the basis of F_0 contour information alone, when segmental information has been eliminated. Certain differences in the F_0 contours of female versus male speakers have been observed (Brend, 1972; Nakatani and Aston, 1980).

Results of a number of studies had suggested that less fundamental frequency variation occurs with increased rate (Lee, 1956; Koshikawa, 1962; Bolinger, 1965; Lightfoot, 1970; Stockwell, 1971). When O'Shaughnessy (1976) measured F_0 change for a single talker speaking at different rates, he found that F_0 rate change or slope was altered rather than the amount of variation. Whether a speaker sacrifices variation or increases rate of F_0 change when talking faster may be idiosyncratic and related to naturalness of the speaking situation.

<u>Perceptual relevance of changes in F_0</u>. Klatt (1973) provided some preliminary answers to the question of listener sensitivity to F_0 movement. Using synthetic vowels with both steady and sloping F_0 he found that listeners are extremely sensitive to changes in slope of F_0. He also

found that perception of F_0 is not necessarily based on energy in the region of the F_0 but the relationship of higher frequency components. In general, it seems that the listener uses relative differences and dynamic aspects of F_0 more than absolute values (Hadding-Koch and Studdert-Kennedy, 1964; Lieberman, 1965; Klatt, 1973).

Although declination seems to be universal (or perhaps because it is universal) t'Hart and Cohen (1973) report that listeners are not consciously aware of it. However, through manipulation of peaks in reiterant speech used in perceptual studies Pierrehumbert (1978) has demonstrated that listeners expect it to occur. On the level of localized attributes the Dutch group found that the slope of F_0 is unimportant perceptually. Rather it is the placement within the syllable of the F_0 peak that is most important. They also indicated that voicing interruptions had little effect on perception of F_0.

Collier (1975) has stated explicitly that F_0 is perceived categorically and that much of the fine-grained movement within the contour is integrated into some patterns to which listeners respond as a gestalt. This notion of categorical perception of intonation contours is at least implicit in the F_0 description systems of Lieberman (1967).

At least one study (Gårding and Abramson, 1965) has investigated listeners' consistency in grouping intonation contours and found that it was not done with a great degree of accuracy. Atkinson's (1973) comment that too little is known about perception of intonation to resolve the issue of categorical perception is probably still valid.

Temporal Phenomena

Speaking Rate

Estimation of speaking rate is one frequently used approach to describing timing in speech. Numerous measures have been employed in determination of speaking rates. One technique that has been used extensively is the word per unit time method (Gilbert and Burk, 1969; Lass and Lutz, 1975; Cartwright and Lass, 1976; Grosjean and Lane, 1976). Syllable per unit time has also been used (Goldman-Eisler, 1968). This latter method may be preferable to others because the average phoneme per syllable ratio of 2.7 appears to be less variable than the number of syllables or phonemes per word (Umeda, 1978).

There is, however, an inherent problem in interpreting any of these measures because their use confounds two separate variables--pausing time and duration of linguistic units. Some indirect approaches to rate determination have been devised which eliminate the pausing factor. These include average phoneme duration (Quinn and Umeda, 1978; Umeda, 1978) ratio of total speech time to number of phonemes per sentence (Quinn and Umeda, 1978) and average seconds per word (Goldman-Eisler, 1961b). Since duration and speaking rate appear to be separately controlled (Harris, 1978) comparison of results obtained using different methods should be done cautiously. References in the literature to duration and pausing separately will now be considered.

<u>Hesitation Pauses</u>

Hesitation pauses have been shown to be related to high-level linguistic processes--primarily ones of lexical choice and syntactic style. Because pauses in spontaneous speech are frequently followed by words of low transitional probability they presumably provide speaker planning time. These breaks in the speech stream have also been associated with level of verbal abstraction and degree of emotionality, but not with syntactic complexity (Goldman-Eisler, 1968).

Support for the psychological reality of linguistic units of varying sizes has been found in the placement and length of hesitation pauses (Mercer, 1973). Goldman-Eisler (1972) presented evidence that presence and length of pauses could be used to differentiate between subordinate and relative clauses. Several investigators have argued for the psychological reality of Trager and Smith's (1951) phonemic clause on the basis of placement of pauses within utterances (Boomer, 1965; Dittman and Llewellan, 1967; Suci, 1967).

In an investigation of the grammatical and acoustic correlates of unfilled pauses, Martin (1970) compared measurements of pausal duration with listener judgments of pause location. The results of this investigation indicated that listeners sometimes perceive pauses to be located where no actual break in the speech stream occurs. These non-silence cued perceived pauses sometimes occurred following abnormally elongated syllables, or pseudo-pauses.

Duration

A consideration of how various factors affect segmental duration must be done against the backdrop of intrinsic phoneme duration. Inherent duration of vowel sounds has been found to be correlated with vowel height (House and Fairbanks, 1953; Peterson and Lehiste, 1960; House, 1961). Consonant durations appear to be determined by manner and place of articulation (Fischer-Jorgensen, 1964; Peterson and Lehiste, 1960). There is evidence that some of these intrinsic durational differences result from physiological constraints (Peterson and Lehiste, 1960) while others are learned as part of the phonological system (Lisker, 1974).

As many as nine factors may influence segmental duration (Coker, Umeda and Browman, 1973; Umeda, 1975). The effects of phonetic environment are well documented (Peterson and Lehiste, 1960; House, 1961; Raphael, 1972) as is the perceptual load associated with variation in voicing (Danes, 1960; Raphael, Dorman, Freeman and Liberman, 1975; Lehiste, 1976).

Duration varies with lexical stress and together with F_0 serves as a cue to the stressed/unstressed distinction (Parmenter and Treviño, 1935; Fry, 1958; Morton and Jassem, 1965). The relative contributions of F_0 and duration in lexical stress may vary depending on certain syntactic and semantic highlighting considerations (Nakatani and Aston, 1981). Most of the variation in duration occurs in the unstressed vowel with stressed vowels being up to 50%

longer than their unstressed counterparts (Parmenter and Treviño, 1935).

Syntax conditions duration through phrase-final or prepausal lengthening which may increase syllable duration by up to 30% (Gaitenby, 1965; Klatt, 1976). Oller (1973) has shown that vowels in nonsense word-final syllables are longer than when in other positions. Klatt (1976) suggests that only phrase and clause final lengthening are significant and that Oller's results occurred because of confounding of word final and phrase final effects. These duration differences appear to serve as juncture cues (O'Malley, Kloker and Dara-Abrams, 1973; Lehiste, Olive and Streeter, 1976; Streeter, 1978). Of all the factors that influence segmental duration those that have the greatest effect are voicing of the post-vocalic consonant and phrase final lengthening (Klatt, 1976). There is, however, some question as to the generalizability of these duration differences (Umeda, 1975; Harris and Umeda, 1974).

Comparison of Speaking Modes

Temporal Phenomena

In this context comparisons will be made between spontaneous versus read speech, as well as between and among readings of different kinds of material. The bulk of these comparisons is concerned with speaking rate.

Conventional wisdom would suggest that people speak more slowly when they are reading than when they are speaking spontaneously. Emprical evidence suggests that this is not necessarily the case. Zwicky (1972) and Shockey (1973) have both found that some speakers speak more rapidly when they are talking spontaneously, while other speakers speak faster when they are reading. With both of these investigators' work it is not clear what is actually being measured since the measures they used confounded duration of linguistic segments and duration of pauses.

In two studies concerned with intelligibility of speech produced at different speaking rates Pollack and Pickett, (1964) found that "slow" conversation was almost as rapid as "fast" reading. This suggests that while there is overlap in the ranges of rates of spontaneous speech and reading, spontaneity in general results in greater speaking rates.

Goldman-Eisler (1961b) compared speaking rates for different degrees of spontaneity in unplanned descriptions and interpretations of captionless cartoons. She found that degree of spontaneity and habit strength affected speaking rate, while level of verbal planning or propositional value did not. Speaking rate was shown to have an inverse relationship with degree of spontaneity, but a direct relationship with habit strength. In these investigations pausal time was carefully defined and shown to be the primary factor affecting speaking rate.

The range of rate variations in conversational speech may be on the average of only 4.4 to 5.9 syllables per second (Goldman-Eisler, 1956) or about 264-354 words per minute. This range of rates for conversational speech appears to be significantly greater than that obtained for reading of the Rainbow Passage (Fairbanks, 1940), 167 to 226 words per minute (Gilbert and Burk, 1969).

Most of the studies on speaking rate have not compared read and spontaneous speech, but rather have been concerned with different reading modes. If degree of carefulness within read speech modes is ordered, certain inferences can be drawn concerning how the variables under consideration might change with spontaneous speech which is presumably less careful than the least careful read speech.

Significant variation in speaking rate has been found to be associated with material read in different contexts. Molter (1974) observed increased rate of selected words when sentences were read in narrative contexts compared to when they appeared either in unrelated lists of sentences or embedded in carrier phrases. Identical sentences appear to be read more slowly when in random list order than when read in narrative context (Umeda, 1978).

Relative duration of speech segments may change with structure of the speaking situation. Vowel duration had been found to vary significantly with the number of syllables per word (Lehiste, 1970; Barnwell, 1971; Klatt, 1973; Lindblom, 1963) when measurements were made of nonsense words, lists of words, and words in short phrases or carrier phrase sentences. This vowel duration dependency on number of syllables per word appears not to function in the connected context mode (Umeda, 1972; Harris and Umeda, 1974). Harris and Umeda have suggested that prosody may play a different role in various speaking modes.

A salient feature associated with spontaneous speech is the presence of hesitation phenomena: repetitions, false starts, filled pauses and unfilled pauses (Macklay and Osgood, 1959). Unfilled pauses are more abundant, of longer duration and occur more frequently within phrase boundaries for spontaneously produced speech samples (Goldman-Eisler, 1968).

Low-level Phonetic and Phonological Differences

Shockey (1974) examined phonological and phonetic reductions in lexically identical read and spontaneous utterances. She found slight quantitative differences in low-level processes between the two speaking styles, but no qualitative ones. Some of the properties that were found to be more heavily represented in spontaneous speech were schwa deletion, monophthongization, vowel centralization and flap deletion. Zwicky (1972) investigated phonological characteristics of casual speech and assumed that the reductions were a consequence of greater speed as a function of decreased formality.

Lindblom (1963) came to a similar conclusion regarding vowel duration. He suggested that reduction of acoustic vowel space was a function of greater speaking rate. While speaking rate is probably one of the variables that affects vowel reduction, other factors such as style (Shockey, 1973) and degree of stress (Gay, 1978; Harris, 1978) contribute to this phonetic process.

Labov (1966) investigated the absence of preconsonantal and final /r/, substitutions of /d/ and /t/ for /ð/, and /θ/, and increase in relative height of two vowels. Each of these phonetic changes increased as speech became more casual.

Smith (1978) used a technique similar to Shockey's (1973) approach for eliciting lexically identical spontaneous and read utterances from deaf children. She found that for 80% of her population, the read versions of utterances were more intelligible than their spontaneous counterparts.

Conclusions

There is a large body of literature on F_0 and duration cataloguing the numerous factors which affect these parameters of the speech signal. Most of these investigations have been carried out on read sentences. A few studies have been done comparing temporal phenomena for readings of different kinds of material and comparing spontaneous and read speech. Only Shockey (1973) and Smith (1978) have used lexically identical read and spontaneous utterances. Given the extensive and multifaceted role that F_0 plays in speech, there is clearly a need to investigate its function in distinguishing between lexically identical read and spontaneous utterances.

CHAPTER III

PROCEDURE

Eliciting Read and Spontaneous Speech Samples

In order to make comparisons between read and spontaneous speech it was necessary to elicit lexically identical utterances under both speaking conditions. This was done by recording speakers in unplanned discussion, transcribing this spontaneous speech and having the speakers read their transcribed conversations.

In obtaining the spontaneous utterances it was necessary to compromise between the need for an appropriate acoustical environment and the need for recording conditions that would encourage spontaneity. The speakers had experience in making recordings and appeared able to speak with a reasonable degree of spontaneity and fluency in a recording booth. Originally four speakers were recorded (two male and two female) but one of the female speakers was eliminated because her conversation contained too few of the syntactically simple sentences that were to form the bulk of the corpus.

A topic for discussion was suggested before the recording sessions although speakers were told they could talk about anything, including their feelings about the recording booth and the recording session. The conversations were dialogues either between the investigator and the speaker (CF) or between two speakers (JP and KM). Approximately one-half hour of spontaneous speech was recorded for each of the three speakers.

The conversations were transcribed adhering as closely as possible to the original utterances. Although false starts, repetitions and filled pauses were eliminated and syntax was corrected in the transcriptions, only a minimum amount of change was made in order to make the spontaneous utterances suitable for reading. The transcriptions were in the form of a script for a play. Each speaker's part was marked by his or her initial. Speakers first read the whole transcribed conversation without having seen it, then re-read it twice. They were unaware that only certain sentences were of interest. For the present investigation only sentences from the initial reading were used. Approximately two weeks elapsed between recording of the spontaneous conversation and the read transcription.

Two of the speakers (KM and CF) have lived in northern New Jersey all of their lives. The third (JP) is a native of Brooklyn, N.Y. and appears to use a dialect consistent with that geographical area.

There was a great deal of variability among speaking/reading styles. KM was an extremely articulate speaker who tended to dominate the conversation. His spontaneous speech contained a large number of false starts. These may have been a mechanism he used to retain the conversational ball during verbal planning periods. KM's reading was fluent. When lexically identical spontaneous and read items were compared, they sounded very much alike. His reading sounded relatively spontaneous and not at all stereotypic. KM's reading might be compared to that of well-trained newscasters who have learned to avoid a blatantly obvious reading style.

In comparison to KM, JP was somewhat less fluent and less articulate. He tended to have filled and unfilled pauses, and repetitions in his speech when he was involved in verbal planning. These differences between KM and JP may have partially been a function of KM's greater experience in doing recordings for speech research. JP was a stereotypic reader who sounded like he was reading when he read. Of the three speakers his speech seemed to differ the most between read and spontaneous versions.

CF presented a different picture altogether. Her spontaneous speech sounded somewhat hesitant, more so than it ordinarily does. This was the first time she had been used as a speaker for speech research at Bell Labs. CF's read speech was more expressive than her spontaneous speech. She appeared more formal when speaking spontaneously than she did when she read. This subjective observation may explain some of the trends found in her data.

Description of Corpus

From over 400 utterances generated across speakers, 66 were excised to be used in this study. The utterances were chosen to conform to the following criteria:

1. Were produced with no perceptible background noise, e.g., movement of a chair or interruption by the other speaker
2. Contained no filled pauses, repetitions, false starts, laughing, coughing, etc.
3. Contained no obvious low level phonetic differences between the two versions, e.g., "gonna" vs. "going to"
4. Were statements
5. Were grammatically acceptable for written language

Seven of the 66 utterances were eliminated from use in the perceptual study because of syntactic complexity. Six of these were used as practice items and were included in the acoustic measurements.

Not all of the utterances were complete sentences. These were either one word utterances, phrases or dependent clauses. Lists of the utterances used are in Tables 1, 2 and 3. There are 25 utterances for CF: 19 sentences and 6 sentence fragments. KM had a total of 21 utterances: 18 sentences and 3 fragments. For JP there were 20 utterances: 13 sentences and 7 fragments.

It was originally intended that unfilled hesitation pauses would be a controlled variable with half the utterances containing such pauses and the other half not containing them. In this context a hesitation pause is defined as a silent period in excess of 25 cs. Upon close examina-

TABLE 1.--Utterances for Speaker CF

1. This is what frightens men.

[a]2. Feminism.

3. This is just what men are afraid of.

4. Everybody's freedom might be considered a threat to someone else.

[b]5. If you're satisfied with it, then you are developing yourself.

6. I didn't assume that.

7. I have a list of things I won't do.

8. You don't know who contributed how much.

9. That's between you and me and the microphone.

[a]10. But I will start to do that shortly.

[a]11. No.

[b]12. I should be, but I'm not.

13. It has been suggested to me that I spend more time on that.

14. I'll have to make some positive plans.

[a]15. Something that they need.

16. You'll help me do that.

17. That's effective communication.

[a]18. Yeah.

[a]19. Norrie.

20. You said a Tuesday would be good.

21. Next Tuesday's no good.

TABLE 1.--Continued

[b]22. You know there's a Chinese restaurant in Plainfield.

23. I know of a good Italian restaurant in Plainfield.

24. Maybe you should try it in your sneakers first.

25. I specifically asked her about articulation.

a Sentence fragments

b Not included in the perceptual study

TABLE 2.--Utterances for Speaker KM

1. There is no N.Y. written accent.

2. They're losers.

3. They're not as healthy.

4. This is an aptitude test.

5. It measures innate ability.

6. They tend to lose things literally.

---

9. It's almost a psychological thing.

10. Not one of them had to do with teaching.

---

14. That's the mark that you've made it.

15. I can't get that excited about the SAT scores falling.

16. I don't think that's our job.

---

[b]18. Some European systems work that way but I couldn't see it.

19. We've always been cranking out kids who were only marginally literate.

20. I've never seen the room that Norm's in.

21. I've never even seen it.

22. You read the same language.

23. That's always the instinct.

[b]24. No, he couldn't.

[a]25. And the others found jobs where they could be trained.

TABLE 2.--Continued

[a]26.	And you wonder how they're going to make it.
[a]27.	But they're not the ones taking the SATs.

a Sentence fragments

b Not included in perceptual task

TABLE 3.--Utterances for Speaker JP

1. I'm not teaching the brighter kids.

[a]2. Well, first of all, let's look at it from the math angle.

3. They don't have that answer in their head.

[a]4. But I have the average kids.

5. I don't have any of the advance classes.

6. You know he'll be able to divide mentally in his head.

[a]7. But the slower kids can't do it, and they don't want to do it.

[b]8. In a graduate course one professor said something that made so much sense.

9. We can't send everybody to college.

10. He's got a personality that doesn't quit.

11. He can liven up any conversation.

---

[a]14. No.

[a]15. Their lack of ability in just fundamental arithmetic.

16. He'll pick it up.

[b]17. He says, "You know someone's got to be a garbage man."

[a]18. Eddy Vent.

19. He's excellent with his hands.

20. He's a dynamite person.

TABLE 3.--Continued

[a]21.	In the sense that he doesn't sit down and read a book.
22.	I still find that utterly amazing.

a Sentence fragments

b Not included in perceptual study

tion of the speech that was produced it was seen that hesitation pauses rarely occurred with fluently produced syntactically simple sentences. Of the 66 utterances only two (CF8 and KM6) contained pauses. Their durations were 110 cs and 33 cs respectively.

The absence of hesitation pauses might be anticipated, since the range of sentence lengths is only four to seventeen syllables, with a mean of 8.79. If the sentence represents an "apperceptive unity" as has been suggested (Blumenthal, 1970) and hesitation pauses function as planning time for the speaker (Goldman-Eisler, 1968) it is not surprising that hesitation phenomena should be sparse in fluent, relatively short simple sentences. Presumably the simple sentence represents a complete thought, the elements of which are likely to have been generated in a wholistic way rather than pieced together as an associative string.

The primary liability associated with investigation of spontaneous speech is the inability to control the linguistic structure of the utterances that are generated. Certain syntactic structures or tokens of speech segments may be required for a particular analysis. Whether or not these will be available cannot be predicted.

Much of the emphasis of the present study was on the fundamental frequency contour. This acoustic phenomenon is extremely sensitive to minor syntactic variation, lexical choice and semantic emphasis. None of these three variables were controllable to any great extent. In selecting the utterances, every effort was made to keep them as syntactically homogeneous as possible. Given that the utterances had to conform to the five criteria listed above, there was little room for syntactic weeding. Therefore, the utterances of the corpus exhibit considerable syntactic heterogeneity.

A simple structural breakdown of the utterances is provided in the Appendix. There are certain similarities among some of the utterances, but many more differences. The vast majority of them have either a personal or demonstrative pronoun as the subject noun phrase. Frequently this pronoun is contracted with the copula or auxiliary verb. There are as many breakdowns of the main predicate verb phrase as there are sentences. This heterogeneity may account for much of the variance seen in the data.

Computer Facilities

Completion of the present study depended on use of the computer facilities in the Communication and Speech Analysis Department at Bell Laboratories in Murray Hill, N.J. The system consists of two computers, a Digital Equipment Corporation PDP 11/45 (Capacity: 128,000 sixteen bit words) and a System Engineering Laboratories (SEL) 75 (Capacity: 64,000 thirty-two bit words). These two systems are connected by hardware and software and share a single hierarchical file system. The SEL terminal (which is one of four SEL units connected to the PDP 11/45) provides interactive time-shared access to the PDP 11/45 which is the main storage system. Communication with the programs of the SCL is done through a Tektronix Memoscope and a keyboard. A Hewlett-Packard Display Unit provides visual representations that include the acoustic waveform, pitch tracks, and spectral sections.

The SCL software system includes an array of interactive programs for analysis, editing and synthesis of speech. For the present study programs performing the following tasks were utilized:

1. Digitization of the speech wave from audio tape at 10,000 samples per second.
2. Examination and editing of the speech wave. The speech editing program yields a visual representation of the digitized speech wave with a cursor that can be located at any sample point. By placement of the cursor it is possible to define the boundaries of any segment of speech for audio playback and/or permanent storage in a speech file.
3. Automatic calculation of the mean F_0 of the speech signal between any two designated sample points (which differed by 10 msec).
4. Extraction of fourteen parameters of the speech signal. It is possible to obtain F_0 and relative amplitude values; these are presented at ten milli-second intervals. The F_0

values are derived through an F_0 tracking algorithm that averages over two to three pitch periods. These values are then used in a Linear Predictive Coding (LPC) analysis which yields twelve pseudo-area parameters representing spectrum. LPC analysis is a method of representing speech with a greatly reduced amount of information. Instantaneous amplitude values are compared with a number of previous ones; in this case the number is twelve. Differences between an instantaneous amplitude value and earlier amplitude values reflect the cross sectional dimensions of the vocal tract along its length.

5. Synthesis of sentences from the fourteen speech signal parameters. The F_0, amplitude and LPC parameters can be used to generate high quality synthesized speech.
6. Manipulation of speech parameters. There is a program which allows the value of any sample point to be altered for any of the fourteen parameters described above.
7. Generation of audio tapes. Audio tapes can be prepared using a playback routine with any specified format.

Perceptual Experiment

Purpose

Given the need to relate possible acoustic differences between read and spontaneous speech to listener judgments, a perceptual experiment was conducted with three purposes. First of all, the perceptual task was designed to discover to what extent listeners could correctly identify utterances as having been read or produced spontaneously when the utterances were presented individually and out of context. A second purpose of the perceptual experiment was the assessment of the role of F_0 contour information in the S/R distinction. The final purpose of the perceptual study was to obtain information upon which to base a choice of acoustic measurements.

Preparation of Stimuli

In order to assess the contribution of F_0 contour information to the S/R distinction it was necessary to present both read and spontaneous versions prepared under two conditions. The first of these was the unaltered condition (UA) in which the sentences were excised from the dialogue, digitized, submitted for LPC analysis and synthesized, but not otherwise manipulated.

The other condition was the altered one (A). Sentences were prepared for the A condition through use of two of the interactive programs on the computer facility described above. The initial step in generating the A versions was to calculate the average F_0 across voiced segments for both versions of each of the utterances. Then the F_0 contours were flattened at the level of the mean F_0 for each version. This was accomplished through setting the F_0 of the initial and final samples of each voiced segment to the mean F_0 and smoothing the contour between them. Contour movement was eliminated while differences in average F_0 between versions was preserved.

Perceptual Judgments

Participants in the perceptual experiment listened to two repetitions of each utterance in isolation and were required to make two judgments. These judgments were whether they thought the utterance was read from a text or was spoken spontaneously, and how confident they were in their choice. In addition to these quantifiable results, listeners were asked to indicate the basis upon which they made their decisions. Copies of instructions to the listeners and the response sheets are in the Appendix.

Listeners and Experimental Design

Forty sophomore and junior speech majors enrolled in a course in articulatory phonetics at Douglass College served as listeners. The perceptual study was conducted at the beginning of the semester so that these listeners could still be considered phonetically naive. The forty listeners were divided up into four groups of ten subjects each.

Each of the sentences listed in the corpus occurred in four different version/condition combinations: UA/S, A/S, UA/R and A/R. Given a total of 59 sentences across speakers, two versions (R and S) and two conditions (UA and A) there was a total of two hundred thirty-six (236) sentences used in the perceptual study. In order to avoid the effect of prior presentation of a version/condition combination on later occurrences of the same utterance in other version/condition combinations, a Latin square type presentation was used. In this type of experimental schema all possible factorial combinations appear only once. Following is a breakdown of the version/condition/utterance combinations heard by each of the four listener groups.

Utterance Numbers	Version/Conditions Heard by Each Group							
	Group 1		Group 2		Group 3		Group 4	
1-15	UA	S	A	S	A	R	UA	R
16-30	A	S	UA	S	UA	R	A	R
31-45	UA	R	A	R	A	S	UA	S
46-59	A	R	UA	R	UA	S	A	S

This breakdown limited the listeners within each group to hearing each utterance only once, but exposed them to all possible version/condition combinations.

All utterances for a given talker were presented in a block, with presentation of utterances randomized within it. Order of presentation of talkers was varied so that each speaker was presented at least once in first, second and third positions.

Practice

Eight practice utterances were included at the beginning of each talker block. These practice items consisted of two sentences which had been deemed syntactically inappropriate for the perceptual study in each of the four version/condition combinations. Exposing listeners to all four version/condition combinations during the practice session was done in order to familiarize them with the strange sound of the altered versions as well as to allow listeners to normalize to each speaker's S/R differences.

The block of practice utterances was the same for each of the four listening groups. The listeners were not apprised as to the practice status of the introductory items.

Acoustic Measurements

Duration

Three duration measurements were made for each version of all the utterances. The first of these measurements was the overall utterance duration. Through comparison of the duration of the two versions the presence of consistent differences in speaking rate could be observed.

The second duration measurement was the length of the final syllable in the utterance. It was used in conjunction with the overall utterance duration in calculating the Final Syllable Ratio (FSR):

$$\text{FSR} = \frac{\text{duration of final syllable}}{\text{overall utterance duration}}$$

A higher FSR for one version of an utterance indicates that the final syllable for that version was proportionally longer than the final syllable for the other version. This ratio was calculated in order to investigate possible differences in sentence final lengthening effects between read and spontaneous versions.

A third measurement made was of the duration of the two syllables in all bisyllabic words. The Stressed/Unstressed Syllable Duration Ratio (SUR) was then calculated to point up possible differences in the durational effects associated with stress differences. The list of bisyllabic words for each of the three speakers is found in the Appendix.

Inherent in the measurement of syllable durations is the problem of parsing the speech stream, which is not segmentable in any absolute sense. Many of the segmentation decisions faced in obtaining these duration measurements were relatively straightforward. That is, they required defining a boundary between phonemes produced by articulatory maneuvers that yield vastly different spectral effects. Abutting phoneme contexts that produced abrupt spectral changes included switches between fricatives and syllabic nuclei.

For certain other contexts the decisions were more problematic. In these less straightforward cases heuristic strategies were specified in advance to assist in making consistent, although somewhat arbitrary decisions across versions. The specifications for segmentation decisions in these problematic cases are given below.

Abutting stops. When two stop consonants are contiguous the first of them is not ordinarily released. The resulting elimination of the presence of a burst or aspiration for the initial stop yields a common closure period for the two plosives. For purposes of specifying syllable boundaries in the cases of adjacent plosives half of the closure was assigned to each of the conjoined syllables.

Abutting nasals. The sole instance where isolating the final syllable required separating conjoined nasals was CF12 where the two syllables were "I'm not." The phonemes /m/ and /n/ are said to be distinguishable on the basis of differences in the degree of second formant transitions (Fry, 1979). In the present case this difference was not obvious. Therefore, half of the nasalized portion was assigned to each of the two syllables.

Abutting fricatives and affricates. There were four instances where segmentation decision around conjoined fricatives had to be made. In each of these cases (z-f z-h, tʃ s-s and z-h) the dispersion of fricative energy or the presence of the low frequency voicing bar was sufficient for disjunction of the two fricative segments.

Abutting vocalics. There were two instances of abutting vowels (/i-I/ and /u-I /) and four of vowels adjoining semi-vowels. (/ɔ -l/) /l-I/ /r-I/ and /l-I/). In each of these cases the movement of the formants and/or increases or decreases in overall energy reflecting changes in degree of mouth opening were considered sufficient basis upon which to make judgments as to phoneme/syllable boundaries.

Word medial semi-vowel. The final word in CF19 was "Norrie." Half of the /r/ duration was assigned to the first syllable and half to the last.

Fundamental Frequency.

Average fundamental frequency. One of the programs on the SEL system described above yields the average fundamental frequency between any two sample points. This averaging was done across all voiced segments, and a mean for the voiced segments was derived for each utterance.

Fundamental frequency range. The high frequency and low frequency points for the two versions of the sentences were determined through visual examination of the pitch contour. The frequency of each of these points was then read directly from the display screen of the computer system. F_0 range was calculated as the difference between these high and low points.

F_0 Contour Analysis

Problems. The fundamental question being asked in the investigation of pitch contours concerns whether there is a consistent difference between the patterns of F_0 movement for the two versions. Such a query involves both the global declination effect and the more subtle localized attributes. Given the highly subjective nature of judgments about declination and localized attributes, a more desirable approach to pitch contour analysis would make use of somewhat more objectively defined parameters. For the present experiment an attempt was made to capture certain regularities in the pitch contours through examination of pitch peaks. One level in the perception of F_0 contours probably involves attention to the relative height of F_0 peaks. Pierrehumbert (1978) demonstrated that listeners can make judgments based on differences in F_0 peak relationships. Specification and measurement of F_0 peaks is somewhat less subjective than the drawing of declination lines by eye as was done by Maeda (1976).

Before going any further it is important to state an assumption that ostensibly underlies the use of F_0 peak measurements in describing declination. This usage tacitly implies acceptance of a theory that explicates intonation on the basis of F_0 levels rather than F_0 change. Such a theory relegates the local movement of F_0 to the realm of artifact, a result of travel from one level F_0 to another. In the present case the decision to examine F_0 peaks is completely pragmatic and theory-free. Given the current state of knowledge about the appropriate way to characterize the F_0 contour, use of any other methods seems rash.

While the decision to look at F_0 peaks rather than the more elusive components of F_0 curves renders the task of F_0 contour analysis somewhat more objective and manageable it by no means charts a trouble-free course. The difficulty stems from the fact stated above that localized movements of F_0 contours have no representation within an F_0 peak analysis system. The elimination of this kind of information is unfortunate because it may serve to obscure significant factors.

With the possibility of localized movements being a potentially fertile source of information, a concerted effort was made to establish a protocol for examining localized attributes. These efforts ranged from attempting to superimpose Dutch group patterns on the contours, to essaying a quantification of F_0 rises and falls between F_0 peaks. These attempts were futile. It became obvious that due to the exploratory nature of this study, the syntactic heterogeneity of the utterances, and the lack of definitive descriptions of localized attributes that investigation of them in this instance would be premature and inappropriate.

Having acknowledged a primary drawback to F_0 peak analysis, some of the problems directly related to F_0 peak measurements should be enumerated. The first step in an F_0 peak analysis must be to determine the location of the peaks of interest. F_0 peaks are superimposed on a declination line. Since a specification of the slope of this gradually falling F_0 base is elusive, it is frequently difficult to define a significant F_0 peak. Is each point which is preceded by a rise and followed by a fall in F_0 to be considered a peak? Or are only those points with a certain amount of prominence in a visual display to be called peaks? Given a steep baseline declination slope, a perceptually significant peak may be visually less conspicuous than an insignificant one superimposed on a more gradual slope.

The problem of relative visual salience of peaks is further exacerbated by certain segmental effects on the F_0 contour. The most obvious of these segmental effects is the high F_0 associated with voiceless consonants. A visually prominent F_0 peak may be an artifact associated with the voiceless consonant environment.

The assumption is being made that peaks in the F_0 contour relate in some way to sentential stress. Most stressed syllables are associated with a peak in the F_0 contour although all F_0 peaks are not necessarily related to stressed syllables. Given this lack of a one-to-one correspondence between F_0 peaks and stressed syllables, and the problem of salience of F_0 peaks it is obvious that the specification of which peaks are to be measured must be made indirectly. That is, one cannot determine which peaks are important through inspection of the peaks themselves. A perceptual judgment of stressed syllables over the course of the utterance must be made to determine the important peaks. The procedure for stress assignment is described in the following section.

If the validity of using stress judgments is accepted, the oversimplification inherent in doing so must be acknowledged. Although increase in F_0 has been shown to be a primary acoustic correlate of stress, duration can also vary significantly with stress level. The role of duration in stress marking and stress perception may assume greater importance for certain speakers and for certain syntactic contexts (Nakatani and Aston, 1980).

The situation is relatively straight-forward when stress is coded in F_0 change. It becomes problematic when duration assumes a prominent role. Phonetic changes reflected in the spectrum may also serve as cues to stress differences. The potentially negative consequences of ignoring duration also holds for disregarding the phonetic variations. Given that the comparisons being made are within speakers and for identical lexical strings the risks associated with neglecting durational and spectral roles in stress are minimized, although not eliminated.

The difficulties discussed so far are inherent in any investigation of F_0 contours. In addition there are at least two problems specifically related to the present study. The first is that the utterances are so syntactically varied. An attempt was made to choose utterances that were as similar as possible, but the way in which the conversation was generated disallowed control of this variable. The utterances which would also be considered complete sentences are broken down into their constituent elements in the Appendix. The second problem unique to this study is that there are some utterances where the main sentence stresses appear to be in different positions. In this case F_0 peak relationships and placement are confounded.

These are the most obvious problems associated with an F_0 analysis. Having fully acknowledged the hazards involved in such an undertaking the next step is to describe the procedure.

<u>Strategy for locating F_0 peaks</u>. The first step in this pragmatic strategy was to identify the perceptually relevant stressed syllables. This was done through independent judgments by the present investigator and a speech/language pathologist with training in articulatory and linguistic phonetics. The few instances where the two judges were in disagreement were discussed and accord was easily reached. This procedure is admittedly completely subjective. However, when considered from the vantage point of the established alternatives it seems the most feasible.

The syllable associated with the designated stressed syllable was located through visual and auditory examination of the pitch tracks. One of the programs on the computer facility that was used for processing and manipulating these utterances allowed great flexibility in viewing and listening to speech samples. Specified segments of speech of any duration (as small as twenty milliseconds) could be displayed alone on the screen or listened to in isolation. The stressed syllables were isolated through this facility and the highest frequency point within each syllable was taken as the F_0 peak. No attempt was made to locate the midpoint of the vocalic portion of the syllable or even to assure that the peak was associated with the vocalic portion.

<u>Organization of F_0 peak information</u>. After the F_0 peaks were identified, located and measured the data were organized in the following ways:

1. A Relative Peak Height Pattern (RPHP) was derived for each utterance. This was done by designating the highest peak as 1 and assigning consecutively higher numbers to the peaks in declining height order. Differences in the RPHPs between the two versions (S and R) were noted. This relative numbering of F_0 peaks resembles the stress level notation system of Trager and Smith and other American linguists. The present usage, however, is based on instrumental measurements rather than subjective stress judgments.

2. A numerical representation of the relationship between the F_0 range and the F_0 peak movement was obtained through the calculation of the Peak Range Ratio (PRR). The formula for calculating this ratio is

$$PRR = \frac{P_h - P_l}{F_0 \text{ range}}$$

where P_h is the highest peak in Hz and P_l is the lowest peak in Hz. The numerator of this fraction indicates the range within which the peaks moved. The ratio itself reflects how much of the total range was traversed by the peaks. A higher PRR indicates that peak movement covered a larger portion of the total F_0 range. It also can be interpreted as evidence of greater movement of the peaks.

3. The Declination Range Ratio (DRR) was calculated in an attempt to obtain a relatively objective indication of the F_0 drift from the beginning of an utterance to the end. The formula for calculating this ratio is

$$DRR = \frac{P_f - P_l}{F_0 \text{ range}}$$

P_f is the value of the first peak in the utterance while P_l is the value of the last peak. All intervening peaks are ignored. A higher DRR indicates that there was greater downdrift for that utterance than for another utterance with a lower DRR.

CHAPTER IV

RESULTS

Spontaneous/Read Identification Task

The data from the S/R discrimination task are given as percent read judgments in Tables 4, 5 and 6 for each of the four version/condition combinations. Results of the ANOVAs for these data are given in Table 7. All statistical analyses in this study were done individually for the three speakers because each speaker had different sentences. The data are presented graphically in Figures 1, 2 and 3.

Inspection of the data in tabular form and the results of the ANOVAs show significant effects for both conditions and versions. For all three speakers there was a significantly larger proportion of read responses for the A condition than there was for the UA condition. F_0 contour flattening was accompanied by an increase in the number of read responses. This relationship held for both spontaneous and read versions.

The data show a significant effect for versions for KM and JP. Read utterances were judged read a significantly greater number of times than their spontaneous counterparts. CF's data showed a lack of significance for versions.

While the percentages of read judgments are useful in

TABLE 4.--Percent Read Judgments by Conditions and Versions for Speaker CI ($\underline{n}$=22)

Utterance Number	UA S	UA R	A S	A R
1	40	20	90	100
2	80	40	60	70
3	50	0	60	50
4	100	40	60	100
6	90	10	60	60
7	0	10	70	100
8	20	100	40	100
9	10	50	90	90
10	10	0	90	100
11	0	70	60	60
13	0	50	80	90
14	30	30	0	100
15	20	0	80	80
16	10	60	100	90
17	20	0	90	100
18	20	30	40	60
19	60	80	60	70
20	70	20	30	80
21	40	20	30	80

TABLE 4.--Continued

Utterance Number	UA S	UA R	A S	A R
23	50	40	100	80
24	0	20	80	20
25	50	0	100	50
Mean	34.1	31.4	67.3	79.1

Note. Some numbers are missing because the utterances were eliminated during the course of the study.

TABLE 5.--Percent Read Judgments by Conditions and Versions for Speaker KM (n=19)

Utterance Number	UA S	UA R	A S	A R
1	70	60	100	100
2	0	0	60	60
3	0	90	0	0
4	50	0	100	100
5	20	60	30	30
6	0	40	70	70
9	0	60	40	40
10	0	30	80	80
14	60	0	100	100
15	10	60	60	60
16	20	0	60	60
19	10	10	70	70
20	0	50	90	90
21	0	0	50	50
22	30	90	100	100
23	20	60	80	80
25	40	50	80	80
26	0	50	40	40

TABLE 5.--Continued

Utterance Number	UA S	UA R	A S	A R
27	0	70	20	20
Mean	17.4	41.0	64.7	77.4

<u>Note</u>. Some numbers are missing because the utterances were eliminated during the course of the study.

TABLE 6.--Percent Read Judgments by Conditions and Versions for Speaker JP (n=18)

Utterance Number	UA S	UA R	A S	A R
1	40	20	60	80
2	0	40	40	70
3	0	100	80	80
4	20	10	50	50
5	20	80	30	70
6	0	40	60	100
7	0	50	20	100
———				
9	0	40	100	100
10	0	20	70	100
11	80	20	60	70
———				
14	40	30	70	80
15	30	80	100	100
16	60	80	40	70
———				
18	40	40	60	100
19	80	70	70	100
20	100	50	30	20
21	80	20	20	30
22	60	40	80	60
Mean	36.1	46.1	57.8	76.6

Note. Some numbers are missing because the utterances were eliminated during the course of the study.

TABLE 7.--Analysis of Variance: Percent Read Responses on S/R Identification Task for Three Speakers

Speaker	Source of Variation	Sums of Squares	DF	Mean Square	F Ratio
	Conditions (C)	454.54	1	454.54	0.660
	Versions (V)	36,004.54	1	36,004.54	52.317**
CF	C x V	1,163.64	1	1,163.64	1.691
	Within group	57,809.10	84	688.20	
	Total	95,431.82	87		
	Conditions (C)	33,264.47	1	33,264.47	47.531**
	Versions (V)	6,264.47	1	6,264.47	8.951**
KM	C x V	580.27	1	580.27	0.829
	Within group	50,389.48	72	699.85	
	Total	90,498.69	75		
	Conditions (C)	12,272.22	1	12,272.22	23.809**
	Versions (V)	4,700.00	1	4,700.00	9.118**
JP	C x V	1,572.77	1	1,572.77	3.051
	Within group	35,050.00	68	515.44	
	Total	53,594.99	71		

** p .01

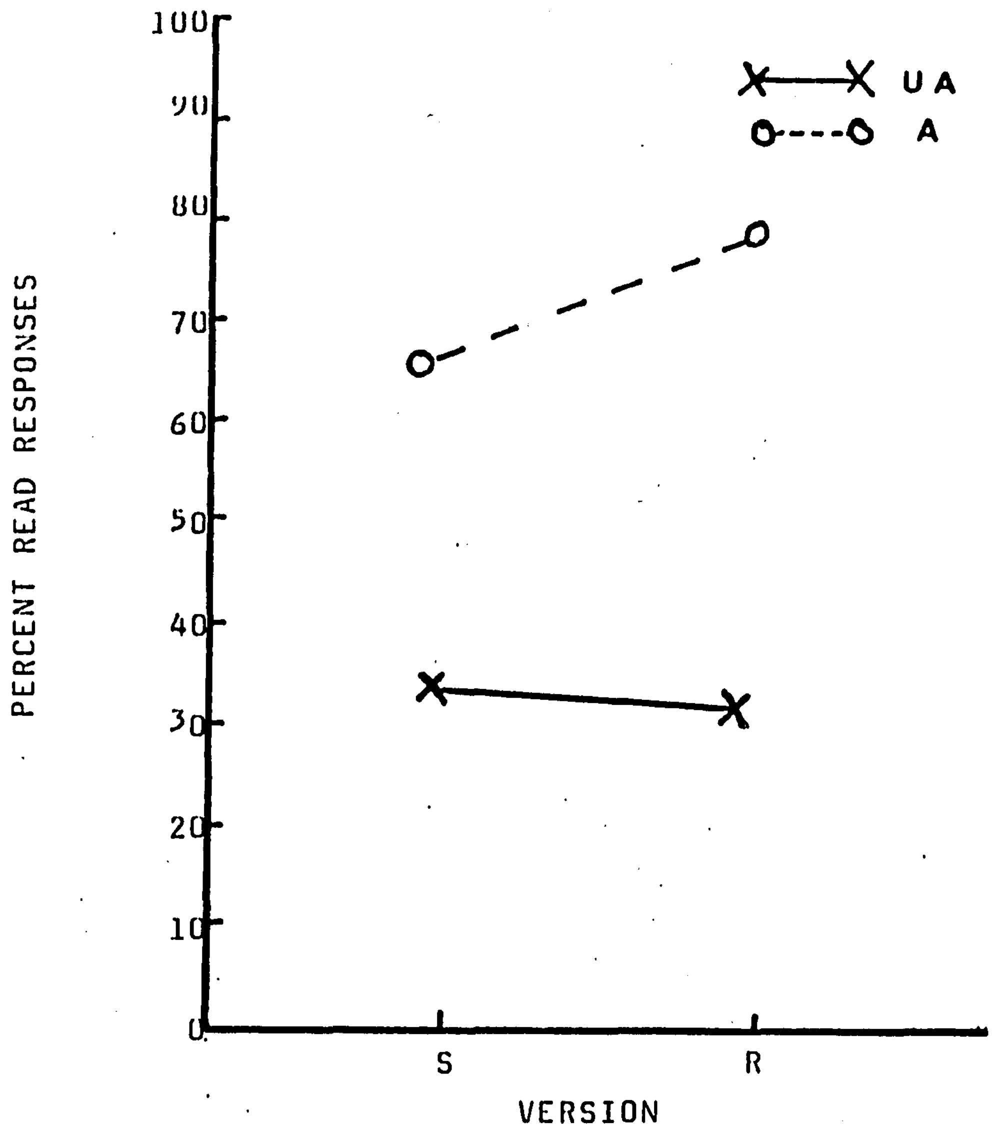


Figure 1. Percent read responses plotted as a function of spontaneous and read versions for unaltered and altered conditions for CF

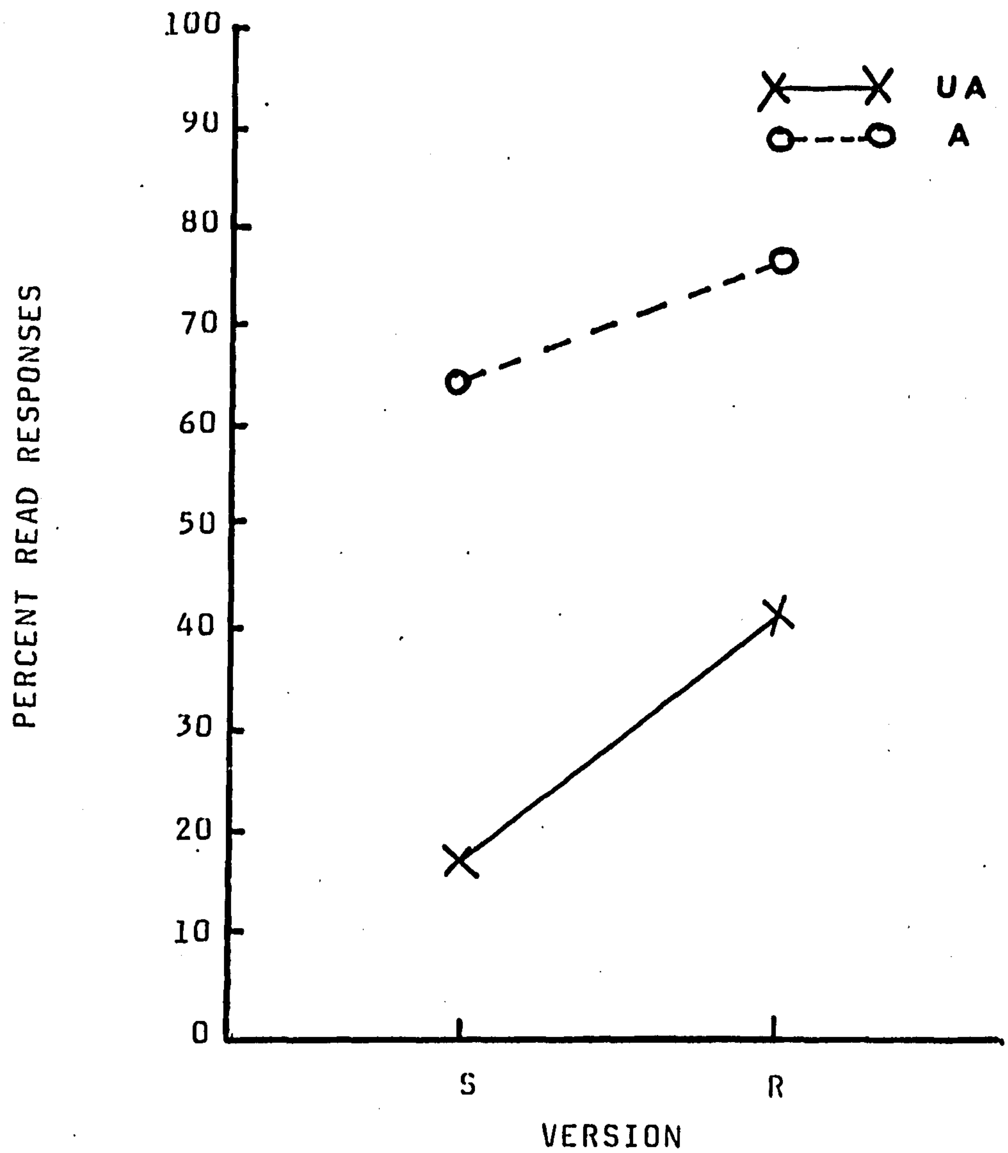


Figure 2. Percent read responses plotted as a function of spontaneous and read versions for unaltered and altered conditions for KM

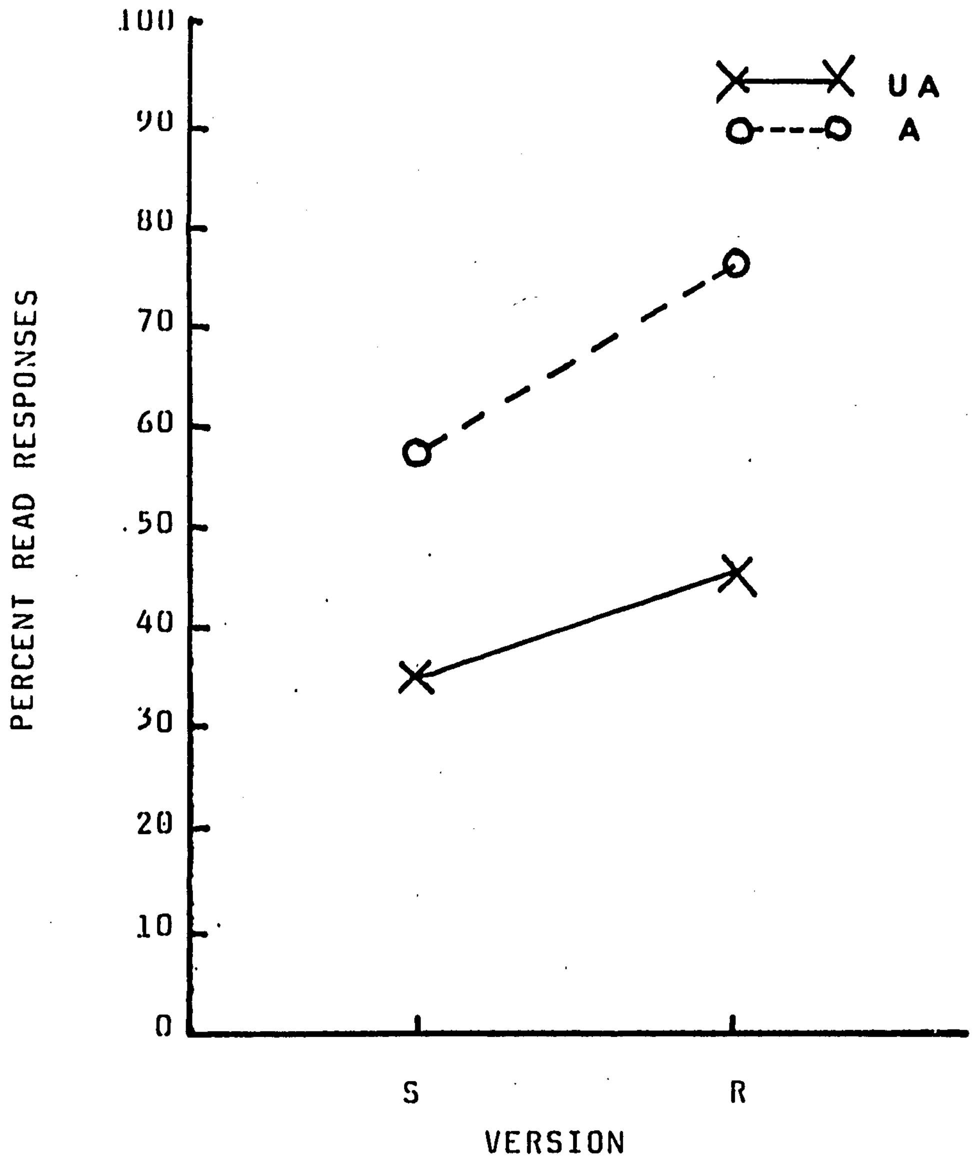


Figure 3. Percent read responses plotted as a function of spontaneous and read versions for unaltered and altered conditions for JP

isolating different response patterns among the four version/condition groups, they do not yield a clearcut answer to the question of the extent to which listeners can make correct S/R identifications. An answer can be obtained through examination of the responses of listeners to the unaltered utterances. In order to evaluate perceptual separation of the lexically identical items, some indication was necessary as to whether a given response score was indicative of chance judgments.

A Chance Judgment Designation (CJD) was assigned to each of the utterances based on the percent of correct responses for the two versions combined. The three assigned CJDs were +, - or 0. Correct responses between 40% and 50% were considered indicative of chance judgments. Utterances with response scores in this range received an 0 CJD. Correct responses of 70% and higher were assumed to represent correct judgments that occurred at a greater than chance level. When the proportion of correct responses for an utterance fell into the 70% to 100% range it was assigned a + CJD. For those utterances with correct responses between 0% and 30% a - CJD was assigned. Percent Correct responses in this low range were assumed to reflect incorrect judgments at a greater than chance level. Although these low-range scores are due to a high number of incorrect responses, they indicate that listeners were responding consistently.

The CJD data for the UA condition are presented in Table 8. Of the 59 utterances used in the perceptual study,

TABLE 8.--Percentage of Utterances Receiving Each Chance Judgment Designation (CJD) by Speakers

CJD	Speaker: CF[a]	KM[b]	JP[c]	Mean
+	23	53	56	42
-	18	10	11	14
0	59	37	33	44

[a] n=22

[b] n=19

[c] n=18

25 of them (42%) were designated as +. This means that somewhat less than half of the utterances could be identified correctly as having been read or produced spontaneously. Forty-four percent of the utterances received an "0" CJD in the UA condition. Eight of the 59 utterances (14%) had - CJDs.

When the numbers of utterances in each of the CJD groups are compared for the three speakers, the data for CF are seen to differ considerably from those for the two male speakers. More than half of the utterances in the UA condition received "+" CJDs for KM (53%) and JP (56%). In contrast, less than a fourth (23%) of CF's utterances received the "+" designation. The greater preponderance of "0" CJDs for CF (59%) as compared with KM (37%) and JP (33%) suggest that CF's utterances were less readily identifiable.

As part of the S/R identification task, listeners were asked to indicate the level of confidence in their judgments. This was done in an attempt to see whether listeners were more confident about their judgments of one or the other versions, or one or the other conditions. The number three was assigned to highest confidence responses; a two to mid-level responses; and a one to the lowest level. Mean confidence levels are presented in Table 9 by type of response (S or R) for the four condition/version categories.

These data show that when listeners respond to unaltered utterances the mean confidence level scores are

TABLE 9.--Mean Confidence Level Scores from Spontaneous/Read Identification Task by Conditions, Versions and Speakers

Listener response	UA/S		UA/R		A/S		A/R	
	S	R	S	R	S	R	S	R
Speaker								
CF	2.42	1.91	2.43	2.10	2.17	2.05	1.91	2.19
KM	2.53	1.99	2.37	1.92	2.04	2.13	1.98	2.17
JP	2.50	2.22	2.13	1.88	1.88	2.23	1.98	2.19
Mean	2.48	2.04	2.31	1.97	2.03	2.14	1.96	2.18

(Column group heading: Condition/Version Combination) higher when the response they give is "spontaneous." The mean confidence level scores across speakers for spontaneous responses to spontaneous utterances (2.48) is higher than that for read responses (2.04). The mean confidence level scores for spontaneous responses to read utterances (2.31) is higher than that for the read response (1.97). Listeners seem more confident when they give a spontaneous response to UA utterances.

Exactly the opposite relationship holds for the altered mean confidence level scores. Read responses have higher scores. For the spontaneous utterances the read mean confidence level score (2.14) is higher than the spontaneous one (2.03). Likewise, for the read versions the read mean confidence level score (2.18) is higher than the spontaneous one (1.96).

This trend for the mean scores across speakers also holds up when each speaker is considered separately, with the exception of CF's altered condition. In this case the spontaneous response to the spontaneous utterances (2.17) was higher than the read response (2.05). Again CF's data are seen to depart from the pattern of the other two speakers.

The confidence level score results show no tendency for the listeners to be more confident in making judgments for one or the other of the versions. Mean confidence level scores reflected the condition. For the UA condition, spontaneous responses were rated more confidently; the opposite was true for the A utterances; the read were rated with greater confidence.

Acoustic Measurements

An attempt was made to identify consistent acoustic differences between the two versions. Measurements were analyzed for all the utterances without regard to their status on the perceptual test. The acoustic measurement data by versions are presented in Tables 10 through 20. In addition, the ranges of data points for each acoustic variable are plotted in Figures 4, 5 and 6.

Duration Measurements

Three duration measurements were calculated. These were (1) overall duration of each utterance, (2) duration of the final syllable in the utterances and (3) the duration of both syllables in each bisyllabic word. Two other indices were derived from the last two measurements.

Overall duration. Overall duration measurements are given in Table 10. The data are summarized in the form of mean durations for the two versions as well as the number of sentences for each version that had the longer duration.

For CF the mean overall duration for spontaneous and read versions was 156 cs and 152 cs respectively. For KM these figures were 151 cs and 143 cs; for JP 182 cs and 183 cs. None of these differences was significant. There is also very little difference between the number of longer utterances within each version category. Of 25 utterances produced by CF, 13 were longer in the spontaneous version and twelve were longer in the read version. The numbers in

TABLE 10.--Overall Duration Measurements (in centiseconds) by Versions and Speakers

	Speaker							
	CF			KM			JP	
Version Utterance Number	S	R	Utterance Number	S	R	Utterance Number	S	R
1	142	131	1	137	135	1	202	151
2	72	55	2	80	74	2	220	281
3	162	146	3	81	78	3	158	182
4	371	175	4	130	119	4	161	131
5	310	311	5	134	118	5	194	187
6	193	92	6	202	161	6	220	297
7	145	156	9	140	149	7	275	285
8	286	181	10	131	156	9	187	190
9	178	174	14	129	118	10	196	196
10	173	165	15	254	256	11	218	191
11	42	27	16	117	119	14	62	41
12	133	106	18	241	243	15	356	325
13	264	288	19	274	275	16	62	67
14	150	175	20	157	154	18	80	60
15	89	93	21	76	90	19	132	149
16	110	89	22	111	104	20	113	125
17	157	141	23	95	112	21	248	246
18	30	21	24	83	77	22	191	185
19	41	51	25	257	185			
20	138	142	26	153	112			
21	104	123	27	181	175			

TABLE 10.--Continued

Version Utterance Number	CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
22	164	201						
23	166	233						
24	181	213						
25	190	223						
Mean	156	152		151	143		182	183
Number longer utterances	13	12		13	8		9	8

each of these categories for KM were spontaneous, 13, read, 8; for JP, spontaneous, 9, read, 8. These results suggest that overall duration was not a variable that differed consistently between the spontaneous and read speaking modes.

The relative dispersion of data points for overall duration can be seen for CF in Figure 4, KM in Figure 5 and JP in Figure 6. CF has a slightly larger range for the spontaneous utterances due to a single extreme data point. For KM and JP the ranges for spontaneous and read utterances are nearly the same.

Final Syllable Ratio (FSR). The ratio of the length of the final syllable to the overall duration (FSR) was calculated for each utterance. The FSRs were included in an attempt to assess whether phrase final lengthening effects differed between the two versions. These data are given in Table 11. There was little difference between the mean FSRs of the two versions for any of the three speakers. CF's mean FSRs were .228 and .208 for spontaneous and read versions respectively. The mean scores for KM were spontaneous, .193 and read, .201. For JP the figures for the spontaneous and read FSRs were .219 and .191 respectively.

The number of sentences for each version which had larger ratios is also indicated. The number of larger FSRs that occurred for each version is nearly the same for KM (S=9; R=11). CF had more larger FSRs for read versions (15)

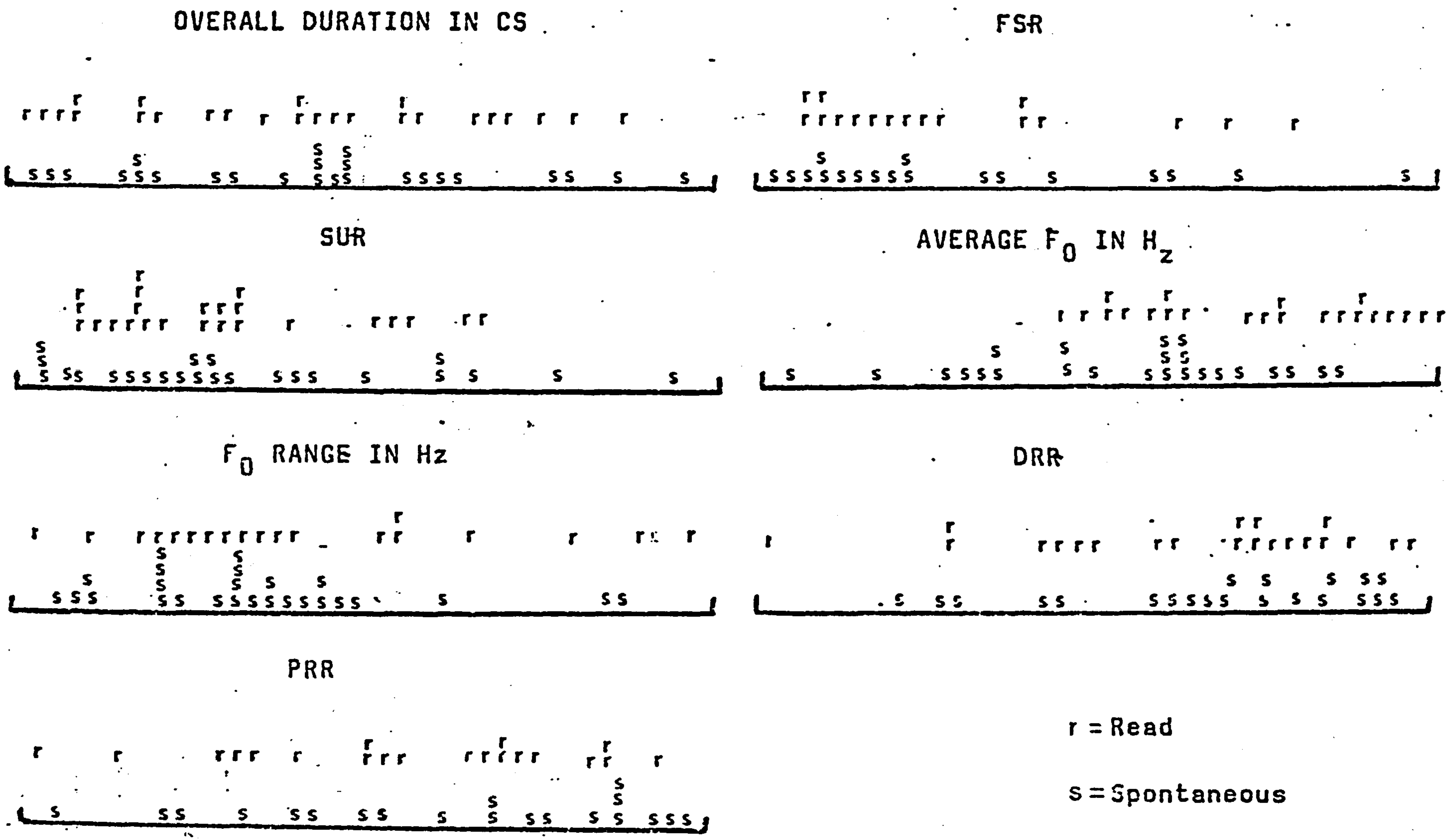


Figure 4. Dispersion of data points for acoustic measurements by versions for CF

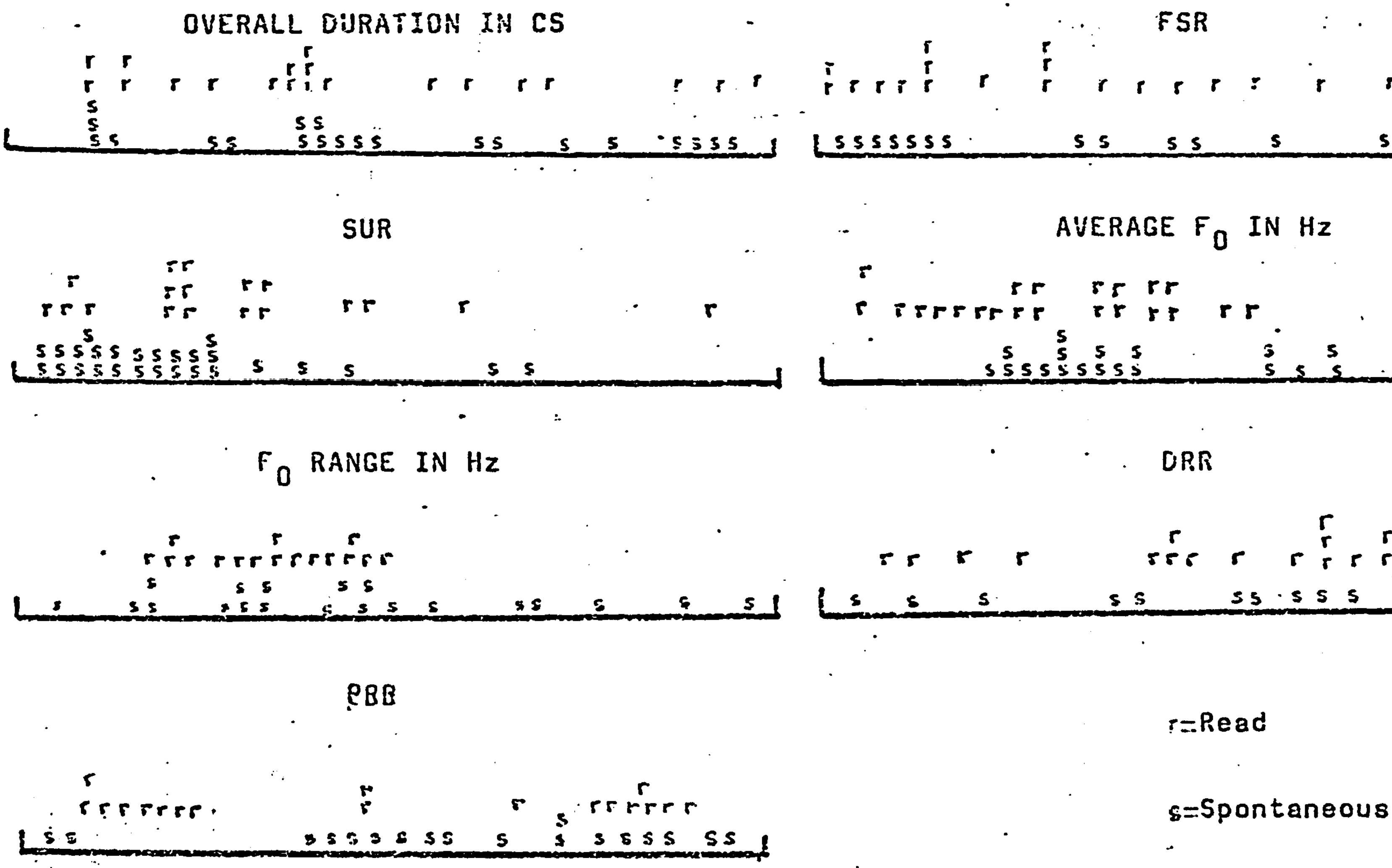


Figure 5. Dispersion of data points for acoustic measurements by versions for KM

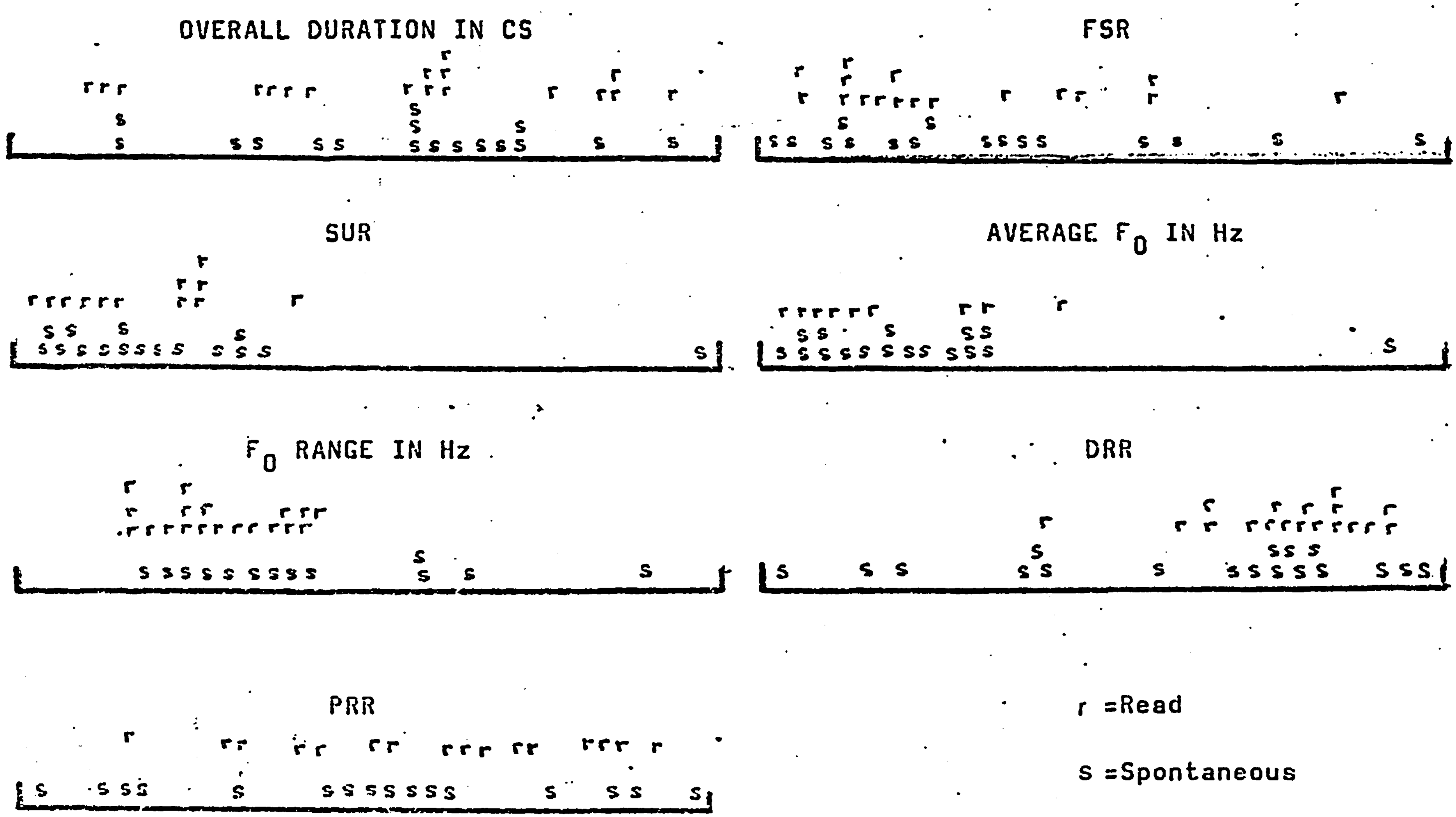


Figure 6. Dispersion of data points for acoustic measurements by versions for JP

TABLE 11.--Final Syllable Ratios (FSRs) by Versions and Speakers

Version Utterance Number	CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
1	.181	.206	1	.131	.126	1	.228	.305
2	.694	.200	2	.487	.446	2	.100	.068
3	.111	.157	3	.568	.500	3	.158	.148
4	.072	.112	4	.223	.286	4	.242	.260
5	.096	.109	5	.119	.119	5	.345	.278
6	.155	.210	6	.129	.130	6	.159	.101
7	.165	.237	9	.207	.215	7	.054	.112
8	.181	.182	10	.145	.109	9	.235	.163
9	.146	.172	14	.077	.110	10	.158	.138
10	.057	.084	15	.094	.078	11	.105	.120
11	*	*	16	.350	.370	14	*	*
12	.278	.330	18	.070	.058	15	.067	.061
13	.079	.087	19	.066	.087	16	.306	.224
14	.440	.314	20	.051	.126	18	.650	.570
15	.393	.473	21	.131	.167	19	.477	.309
16	.390	.303	22	.135	.192	20	.212	.176
17	.229	.156	23	.284	.321	21	.097	.106
18	*	*	24	.277	.260	22	.131	.108
19	.560	.412	25	.222	.243			
20	.174	.218	26	.065	.053			
21	.231	.219	27	.226	.217			

TABLE 11.--Continued

Version	Speaker CF S	CF R		KM S	KM R		JP S	JP R
Utterance Number			Utterance Number			Utterance Number		
22	.195	.149						
23	.060	.146						
24	.254	.164						
25	.095	.136						
Mean	.228	.208		.193	.201		.219	.191
Number larger ratios	8	15		9	11		12	5

* Single syllable utterances

than for spontaneous versions (8). Speaker JP's distribution was the opposite. He had more larger FSRs for his spontaneous versions (12) than for his read versions (5). The x^2s computed for CF's FSRs (x^2=2.13) and JP's FSRs (x^2=2.88) were not significant. The x^2s were not computed for KM's data because of the closeness of the two numbers. Phrase-final lengthening effects, as measured by the FSR appeared not to differ between versions.

From Figures 4, 5 and 6 it can be seen that for each of the three speakers the range of FSR data points is greater for the spontaneous versions. In each case this difference is due to the presence of one or two higher and lower measurements.

Stressed/unstressed ratios. The stressed/unstressed ratio (SUR) for each of the bisyllabic words was calculated in an attempt to assess whether duration as a cue to stressing was used differently between the two versions. These data are summarized in Table 12 as mean SURs for each version as well as the number of larger SURs in each version category.

For CF the mean SURs were 2.51 for spontaneous utterances and 2.11 for read utterances. Of her total sixteen bisyllabic words, eight had larger SURs for the spontaneous versions and eight for the read versions. Speaker KM, with 27 bisyllabic words, had a mean SUR for spontaneous versions

TABLE 12.--Stressed Syllable/Unstressed Syllable Duration Ratios (SUR*s) of Bisyllabic Words for Three Speakers

	Speaker							
	CF			KM			JP	
Version Utterance Containing Word	S	R	Utterance Containing Word	S	R	Utterance Containing Word	S	R
1	1.67	1.57	1	1.24	1.29	1	9.33	2.54
3	5.00	4.47	1	1.14	1.33	1	2.45	3.46
4	0.89	1.06	1	0.92	1.16	2	1.60	2.04
4	2.22	2.38	2	0.59	0.57	3	0.73	1.1
5	1.44	0.74	3	1.00	1.00	4	0.48	1.32
6	12.00	7.75	5	0.72	0.86	5	1.46	1.48
9	2.76	2.35	5	2.56	0.51	5	1.33	2.28
10	3.36	1.52	9	0.33	0.41	6	1.92	2.46
15	3.28	1.85	10	2.27	3.35	7	2.71	2.33
19	0.59	1.12	15	1.67	1.27	8	0.76	0.75
20	1.35	2.11	15	3.75	6.25	8	2.14	1.28
21	1.22	1.33	18	0.85	0.94	10	1.69	2.17
22	0.61	0.67	18	1.00	2.17	11	2.00	1.71
22	2.18	3.37	19	0.44	0.90	11	1.69	1.11
24	0.67	0.75	19	1.45	2.68	18	1.17	1.08
24	0.89	0.65	19	1.35	2.00	20	0.87	1.00
			20	1.23	2.00	21	2.20	1.62
			21	0.80	1.44			
			21	0.95	1.12			
			22	1.34	1.02			

TABLE 12.--Continued

Version Utterance Containing Word	CF S	CF R	Utterance Containing Word	KM S	KM R	Utterance Containing Word	JP S	JP R
			23	0.46	0.42			
			23	1.45	0.66			
			24	1.39	1.05			
			25	0.27	0.29			
			26	3.43	3.00			
			26	2.17	1.71			
			27	1.07	0.94			
Mean	2.51	2.11		1.33	1.49		2.03	1.75
Number of larger SURs	8	8		10	16		8	9
				1 same				

*SUR = Duration of stressed syllable / Duration of unstressed syllable of 1.33 with a mean SUR for read versions of 1.49. Ten of these 27 bisyllabic words had larger SURs for the spontaneous versions, 16 were larger for the read versions, and one had the same SUR for both versions. The SURs for speaker JP's spontaneous and read versions were 2.03 and 1.75 respectively. The corresponding number of larger SURs was eight and nine for the two versions. Duration in terms of lexical stress differences did not vary significantly between versions. Both CF's and JP's SUR data points were more widely dispersed for the spontaneous utterances. KM had a single extreme SUR for one of his read utterances.

There was a total of nine duration measurements (three measures x three speakers). For six of the nine measurements the spontaneous utterances had a wider dispersion of data points. Two of the measurements had the same range for both versions. In only one of the nine cases was the read dispersion greater than that of the spontaneous (KM's SURs). For the duration measurements there was a clear tendency for the range of data points of the spontaneous versions to be wider than those of their read counterparts.

Fundamental Frequency Measurements

In attempting to assess the role of F_0 in the S/R distinction two kinds of variables associated with F_0 were measured. The first kind was made up of two aspects of the F_0 in general: average F_0 and F_0 range. The second type of variable was related to the contour that describes changes in F_0 over time.

Average fundamental frequency. Average fundamental frequencies were obtained for each sentence. Table 13 contains these data for all utterances, the means by version, and the results of t tests. The values for the mean average fundamental frequencies were: CF--S=175, R=188 (t=0.03); KM--S=115, R=103 (t=1.03); and JP--S=113, R=105 (t=0.69). None of these differences was significant.

The number of sentences in each version category that had the higher average fundamental frequency is shown in Table 14. Results of analyses are included. When viewed from this vantage point some significant results appear. For CF it can be seen that twenty-one of her twenty-five utterances had higher fundamental frequencies for their read versions (χ^2=11.56, $p < .01$). For the other speakers this relationship was reversed. Eighteen of KM's 21 utterances had higher fundamental frequencies for the spontaneous versions, χ^2=10.71, $p < .01$. Twelve of JP's 18 spontaneous utterances had higher average fundamental frequencies than their read counterparts, while only two read ones had higher F_0s. The remaining four had identical F_0s for both versions, $\chi^2 = 7.14$, $p < .01$. Thus it can be seen that there were small but consistent differences between versions in average F_0, although the direction of these differences varied with speakers. Again the data for speaker CF differ from those for the other two speakers.

TABLE 13.--Average Fundamental Frequencies by Versions and Speakers

	Speaker							
	CF			KM			JP	
Version Utterance Number	S	R	Utterance Number	S	R	Utterance Number	S	R
1	187	193	1	111	94	1	109	109
2	138	175	2	114	97	2	110	104
3	200	203	3	121	111	3	110	101
4	177	186	4	105	109	4	122	108
5	197	191	5	105	98	5	108	98
6	160	172	6	112	108	6	102	98
7	181	201	9	113	107	7	126	102
8	169	171	10	111	100	9	114	112
9	166	173	14	108	101	10	107	107
10	161	187	15	154	104	11	112	112
11	172	197	16	105	90	14	112	98
12	181	183	18	120	109	15	108	104
13	170	185	19	110	101	16	106	102
14	166	178	20	108	101	18	103	106
15	180	174	21	108	110	19	122	109
16	166	170	22	122	90	20	147	100
17	147	196	23	102	109	21	105	109
18	188	217	24	110	107	22	103	103
19	182	208	25	120	103			
20	191	177	26	131	117			

TABLE 13.--Continued

	Speaker							
	CF			KM			JP	
Version Utterance Number	S	R	Utterance Number	S	R	Utterance Number	S	R
21	180	177	27	122	105			
22	193	194						
23	176	180						
24	182	212						
25	175	199						
Mean	175	188		115	103		113	105

TABLE 14.--Number of Utterances having Higher Average Fundamental Frequencies within Version Categories by Speakers (Results of Chi-square Analyses Included)

	Speakers		
Versions	CF	KM	JP
Spontaneous	4	18	12
Read	21	3	2
Chi-square	11.56**	10.71**	7.14**

** p .01

When the dispersion of F_0 data points in Figures 4, 5 and 6 are examined, it can be seen that the spontaneous range is wider than the read range. In this case the extension of range was at one end of the continuum. The average F_0 for the spontaneous versions was lower for CF and the extreme data points were in the low F_0 region. In contrast, the extreme data points for KM and JP, whose spontaneous versions tended to have higher average F_0s, were at the high end of the frequency range.

Fundamental frequency range. The F_0 range was calculated for each sentence as the difference in Hz between the highest and the lowest occurring frequencies. These data are presented in Table 15. There was no significant difference between the mean ranges for the two versions for any of the speakers. The mean F_0 range for CF's spontaneous versions (94Hz) was lower than the mean range for her read versions (106 Hz), but this difference was not significant, T (24) =0.15. There was also little difference between the number of larger ranges for CF's two versions (S=10; R=14). The mean scores for the other two speakers went in the opposite direction: Spontaneous mean ranges were higher than read mean ranges. The figures for KM were--S=62, R=54, t (20) = 0.0693; for JP--S=70, R=59, t (16)=0.439. The differences between the two versions in distributions of the number of larger F_0 ranges was small for KM (S=12; R=8) and non-existent for JP (S=10; R=10). F_0 range ap-

TABLE 15.--Fundamental Frequency Ranges by Versions and Speakers

Version Utterance Number	CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
1	127	129	1	21	49	1	46	74
2	80	70	2	69	54	2	62	88
3	142	252	3	35	56	3	40	58
4	115	138	4	44	81	4	42	54
5	244	239	5	58	34	5	38	46
6	59	187	6	47	30	6	49	41
7	96	166	9	49	44	7	156	62
8	79	62	10	57	36	8	48	75
9	97	54	14	35	39	9	79	72
10	42	113	15	143	71	10	70	62
11	31	19	16	50	71	11	40	50
12	81	96	18	85	85	14	25	26
13	89	131	19	84	72	15	129	69
14	60	63	20	36	46	16	32	34
15	59	70	21	59	60	17	129	84
16	71	92	22	46	41	18	32	52
17	58	131	23	49	53	19	92	82
18	47	31	24	95	61	20	157	46
19	43	78	25	61	42	21	51	43
20	77	76	26	60	51	22	79	55

TABLE 15.--Continued

Version Utterance Number	CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
21	62	51	27	125	51			
22	156	84						
23	102	107						
24	247	210						
25	82	105						
Mean	94	106		62	54		70	59
Number larger ranges	10	14		12	8		10	10

peared not to vary consistently between the read and spontaneous utterances examined in this study.

Both KM and JP had greater dispersion with their spontaneous F_0 range data points. For KM's measurements both extra high and extra low extreme points occurred. JP's larger range was the result of several utterances which had extremely large F_0 ranges. The dispersion relationship for CF's F_0 range data are opposite from that of the two male speakers. There was greater dispersion for the read data points than for the spontaneous ones.

Relative Peak Height Pattern. The RPHP was determined by assigning the number one to the highest peak within an utterance with decreasingly lower peaks receiving progressively lower numbers. The RPHPs for all of the utterances are presented in Table 16. This information is summarized in terms of the number of utterances that have the same RPHP for both versions. For each of the speakers, it can be seen that less than half of the utterances had exactly the same pattern for the two versions. CF had a higher portion of her sentences the same (9 out of 25: 36%) than did JP (2 out of 20: 10%). This suggests that JP's RPHPs were more likely to change between versions. KM's tally was intermediate: 6 of his 21 utterances (29%) have the same pattern.

Table 16 also includes the number of utterances which had perceptually relevant peaks on different syllables for the two versions. Of CF's 25 utterances, 7 differed in this

TABLE 16.--Relative Peak Height Patterns (RPHP) by Versions and Speakers

Version Utterance Number	CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
1	321	123	1	312	231	1	12	123
2	1	1	2	1	1	2	1234	123
3	213	231	3	21	21	3	132	122
4	1222	1234	4	11	12	4	12	12
5	132	132	5	21	21	5	11	12354
6	123	123	6	21	11	6	1243	123
7	123	123	9	11	1	7	312	123
8	213	123	10	12	12	8	2132	1234
9	231	231	14	121	213	9	231	213
10	12	12	15	123	132	10	132	2143
11	1	1	16	12	12	11	213	121
12	12	12	18	321	321	14	1	1
13	12	132	19	132	12	15	3241	1243
14	1212	1111	20	1243	3124	16	21	12
15	21	11	21	12	21	17	132	142
16	123	12	22	12	21	18	1	1
17	12	21	23	21	12	19	12	12
18	1	1	24	21	21	20	213	121
19	1	1	25	132	131	21	321	1243
20	123	112	26	112	1	22	231	123

TABLE 16.--Continued

Version Utterance Number	Speaker: CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
21	21	21	27	12	21			
22	123	213						
23	213	4123						
24	1234	1234						
25	12	12						
n	25			21			20	
Number of utterances with same RPHP for both versions	9			6			2	
Number of utterances with same slope for both versions	15			10			7	
Number of utterances with different peak positions between versions	7			3			8	

way, as did 3 of KM's 21 utterances and 8 of JP's 20 utterances.

The RPHP was included in an attempt to capture any similarities and/or differences in pattern of the topline that might exist between versions. Given the relatively small corpus size and the large number of different patterns that emerged, the usefulness of the RPHP is limited. One thing that is revealed, however, is that the corpus used for this study contains a number of utterances whose toplines differ markedly from those seen in the writings on the F_0 contour. Many of these utterances do not exhibit topline declination. A number of them actually have final peaks which are higher than initial ones. Only 34 of the 132 utterances across speakers and versions had a 12, 123 or 1234 RPHP. These three RPHPs would be the ones associated with typical declining contours.

<u>Slope direction</u>. In order to capture the initial peak/final peak relationship an arrow description of the slope was derived from the relationship of the first and last peaks of each utterance. Any intervening peaks were ignored. If the final peak was within ± 5 Hz of the F_0 of the initial peak the utterance slope was designated as flat (⟶). For utterances whose final peaks were more than 5 Hz higher than initial peaks, the utterance slope was designated as rising (↑). When the final peak was more than 5 Hz lower than the initial peak the utterance

received a falling slope designation (↓). While capturing a certain regularity about the slopes of the utterances these designations eliminate a lot of information as to relative differences in slopes of utterances which may fall into the same slope categories. The Declination Range Ratio, discussed in the next section, yields more information of a relative nature.

Table 17 contains a breakdown of the number and percentages of slope types by version and speaker and for versions combined by speakers. There were 17 rising contours among the spontaneous utterances and 13 among the read; 37 falling for the spontaneous and 38 among the read; 5 flat contours among the spontaneous utterances and 6 among the read. These differences were too close to warrant statistical treatment.

When the numbers of slopes in each category between versions are considered by individual speakers there is one trend worth noting. JP has six spontaneous rising contours, but no read ones. The number of his falling contours increases from 10 for the spontaneous utterances to 16 for the read ones. This suggests that for JP declination may be more fully realized with read utterances.

The data in Table 17 show that a substantial number of utterances in this natural, conversational mode had atypical rising or flat F_0 contours. Most of the F_0 contours seen in the literature have falling slopes. Of the 116 utterances across versions, 75 (65%) had the downward

TABLE 17.--Number of Utterances in each Slope Category by Versions and Speakers; Percentage of Utterances in each Slope Category across Versions by Speakers; Percentage of Utterances in each Slope Category across Versions and Speakers

Version	S			R			Percent of S+R[a]			
Speaker	CF	KM	JP	CF	KM	JP	CF	KM	JP	Total[a]
Slope Category	4	7	6	5	8	0	21	40	17	26
	17	10	10	14	8	16	74	47	72	65
	0	3	2	2	2	2	5	12	11	9

[a]116 utterances across versions and speakers

sloping contour. Thirty (26%) had rising contours. The remaining 11 (9%) had a flat slope. A total of 41 utterances (35%) had atypical contours.

There is a difference in the pattern of distribution of slope categories among speakers. KM had proportionally more atypical slopes than did the other two speakers. Only 47% of his 42 utterances (combined across versions) declined. This is in comparison with 74% for CF and 72% for JP.

In addition to the number of occurrences of the three slope categories, changes between versions from one slope to another were considered. These data are presented in Table 18. There are two questions to be answered from these data. The first of these questions has to do with the proportion of utterances which have slope changes between versions. Of the 66 utterances across speakers, 24 of them (36%) had slope changes. When speakers are considered individually it can be seen that CF had the smallest number of utterances which changed slope between versions--24%. KM had 38% slope changes and JP had 50%. The greater percentage of slope changes for JP suggests that slope may play a bigger role in S/R identification of his utterances than in those of the other two speakers.

The data in Table 18 are also broken down into the six possible categories of slope change from one version to the other. The changes of greatest interest are those that go from rising to falling (↑ ↓), and from flat to falling,

TABLE 18.--Number of Utterances in all Slope Change Categories by Speakers; Percent of Total Utterances across Slope Change Categories by Speakers

Slope Change Category (S-R)	Speaker CF	KM	JP	Total
↑↓	1	2	6	9
→↓	0	2	2	4
↓↑	3	2	0	5
↓→	1	1	2	4
↑→	1	1	0	2
→↑	0	0	0	0
Total	6	8	10	24
Percent of Total Utterances	24	38	50	41

(→ ↓). Such patterns suggest that declination is more strongly represented in the read versions. The second question that these data can answer is what proportion of the utterances fall into these two categories. Of the 24 utterances that have slope changes between versions, 13 of them went from rising or flat to falling. Across speakers there is no clear trend for declination to be better represented in read utterances.

There is a different pattern in slope changes for the three speakers. Of JP's 10 utterances which changed between versions, eight have either rising or flat contours for the spontaneous versions, and falling contours for the read ones. This is in contrast to four for KM and one for CF. These differences between speakers suggest that declination may be more strongly represented in JP's read versions in comparison with the other two speakers.

Declination Range Ratio. Although the RPHP and slope directions provide information for categorizing F_0 contours, they require supplementation by a quantitative index which will be sensitive to slope differences within categories. The Declination Range Ratio (DRR) reflects movement of the topline as defined by the F_0 peaks. DRRs for all sentences, with the exception of those having a single peak, are given in Table 19. Examination of these data show that for speakers CF and KM there is no difference between the versions on the basis of the DRR. For speaker CF the mean DRRs for the spontaneous and read versions were .723 and .692 respectively.

TABLE 19.--Declination Range Ratios (DRRs)[a] by Versions and Speakers

Version Utterance Number	CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
1	.730	.755	1	.750	.611	1	.738	.759
2	___[b]	___[b]	2	___[b]	___[b]	2	.784	.771
3	.380	.963	3	.563	.385	3	.757	.835
4	.781	.811	4	.694	.597	4	.753	.747
5	.337	.377	5	.380	.709	5	.739	.794
6	949	.846	6	.578	.677	6	.765	.851
7	.816	.680	9	.866	.944	7	.062	.839
8	.721	.747	10	.745	.779	8	.676	.811
9	.269	.630	14	.733	.734	9	.492	.756
10	.690	.798	15	1.000	1.000	10	.629	.636
11	___[b]	___[b]	16	.813	.608	11	.721	.680
12	.961	.842	18	.308	.313	14	___[b]	___[b]
13	.667	.554	19	.723	.796	15	.139	.775
14	.572	.620	20	.868	.856	16	.644	.925
15	.547	.588	21	.723	.587	17	.984	.516
16	.731	.703	22	.746	.794	18	___[b]	___[b]
17	.791	.004	23	.711	.758	19	.755	.753
18	___[b]	___[b]	24	.246	.447	20	.833	.671
19	___[b]	___[b]	25	.745	.711	21	.530	.916
20	.909	.897	26	.731	.722	22	.486	.897

TABLE 19.--Continued

Version Utterance Number	Speaker CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
21	.963	.761	27	.693	.596			
22	.945	.766						
23	.668	.372						
24	.876	.935						
25	.879	.878						
Mean	.723	.692		.719	.681		.638	.774
Number of larger DRRs	10	10		9	9		5	12
Number of same DRRs	1			2			1	

[a] $DRR = \frac{F_0 \text{ 1st peak} - F_0 \text{ last peak})+100}{F_0 \text{ highest peak} - F_0 \text{ lowest peak}}$

[b] Utterances having a single peak

Of the 21 utterances for which the DRR was computed, 10 had higher DRRs for the spontaneous versions, ten had higher DRRs for the read versions, while one was the same for both. These values for KM were .719 for spontaneous versions, .681 for read versions. He had nine higher DRRs for each version, with two the same.

The mean DRR for JP's read versions was .774 in comparison to a mean of .638 for the spontaneous versions. This difference was significant at the .05 level (t,36 df = 2.13). These data suggest that topline declination for JP's read utterances is often greater than that for his spontaneous ones. While the slope directions were able to pick up some differences between versions in JP's utterances, examination of the DRRs showed that topline might differ between versions that were in the same slope category.

The dispersion data in Figures 4, 5 and 6 for the DRR indicate that both KM and JP had larger ranges for the spontaneous versions. This difference in range is small for KM, extended by a single data point at the low end of the continuum. The difference in ranges between spontaneous and read for JP is more pronounced, with four DRR data points considerably lower than the lowest read DRR. CF's DRR data depart from the pattern of the two male speakers in that her read DRR dispersion is greater than that of her spontaneous.

Peak Range Ratio. Spontaneous speech is frequently characterized as being more "expressive." One of the listeners in this study included this as a criterion by which she made her S/R decisions. One of the acoustic factors that could contribute to this subjective notion of greater expressiveness is movement of the F_0 contour. Such movement should be reflected in the range of peak movement.

The Peak Range Ratio (PRR) was calculated in an attempt to show how much of the F_0 range was covered by movement of the peaks. These data are presented in Table 20 for all of the utterances across speakers. Examination of the data reveals that there is a significant difference in PRRs between read and spontaneous versions for only one of the speakers--KM. The mean PRR for CF's spontaneous versions (.600) was higher than that for the read ones (.523) but this difference was not significant. Twelve (12) of her spontaneous versions had higher PRRs as compared with eight read versions which had higher PRRs. This difference was also insignificant.

For speaker KM the mean PRRs for the spontaneous and read versions were respectively, .438 and .312, $\underline{t}\,(38)=4.535$, $\underline{p} < .01$. There were seventeen utterances whose spontaneous versions tended to have significantly higher PRRs than their read counterparts, $x^2(1) = 9.8$, $\underline{p} < .01$.

For speaker JP there is not a significant difference between the means for the two version categories (S=.429; R=.488). The direction of this difference is the opposite from that found in KM's data. Five of the utterances had higher PRRs for their spontaneous versions, while thirteen

TABLE 20.--Peak Range Ratios (PRR) by Versions and Speakers

	Speaker							
	CF			KM			JP	
Version Utterance Number	S	R	Utterance Number	S	R	Utterance Number	S	R
1	.722	.643	1	.904	.306	1	.152	.432
2	—[a]	—[a]	2	—[a]	—[a]	2	.435	.511
3	—[a]	—[a]	3	.656	.714	3	.450	.552
4	.591	.634	4	.023	.099	4	.167	.278
5	.688	.791	5	.689	.147	5	.053	.348
6	.480	.453	6	.170	.033	6	.408	.488
7	.625	.488	9	.024	.523	7	.756	.581
8	.418	.338	10	.298	.167	8	.125	.560
9	.639	.148	14	.314	.154	9	.430	.514
10	.047	.619	15	.818	.704	10	.771	.371
11	—[a]	—[a]	16	.440	.056	11	.400	.240
12	.913	.677	18	.506	.494	14	—[a]	—[a]
13	.292	.305	19	.536	.514	15	.612	.681
14	.504	.510	20	.694	.543	16	.469	.706
15	.220	.014	21	.254	.100	17	.318	.476
16	.352	.380	22	.196	.073	18	—[a]	—[a]
17	.431	.756	23	.408	.302	19	.487	.451
18	—[a]	—[a]	24	.547	.459	20	.885	.152
19	—[a]	—[a]	25	.574	.476	21	.392	.744
20	.792	.763	26	.233	.176	22	.418	.709

TABLE 20.--Continued

Version Utterance Number	CF S	CF R	Utterance Number	KM S	KM R	Utterance Number	JP S	JP R
21	.903	.294	27	.448	.196			
22	.910	.631						
23	.823	.476						
24	.826	.905						
25	.822	.635						
Mean	.600	.523		.438	.312		.429	.488
Number of higher PRRs	12	8		17	3		5	13

a Utterances having a single peak

had higher PRRs for their read versions. This difference was not significant, χ^2 (1) = 3.55, although an χ^2 of 3.841 would have been significant at the .05 probability level.

Data point dispersions of PRR in Figures 4, 5 and 6 show larger ranges for the spontaneous versions for KM and JP. The ranges were extended by one or two points on either end of the continuum. CF's ranges of PRR data points for spontaneous and read versions were essentially the same.

It should be noted that although the PRR does reflect greater peak movement, it is insensitive to the difference in movement between adjacent peaks, and the distance between peaks and valleys. Therefore, similar PRRs do not necessarily preclude greater peak movement for one of the two utterances being compared. Such a difference might be found if another index of peak movement were used.

CHAPTER V

DISCUSSION

Spontaneous/Read Identification

The initial consideration in this investigation was the extent to which listeners could differentiate between lexically identical spontaneous and read utterances. An indirect approach to assessing this ability is through consideration of the percent read judgments for the two versions. Figures 7, 8 and 9 are based on the same percent read data as Figures 1, 2 and 3, but they have been replotted to highlight the version contrast. These figures clearly illustrate the greater percentage of read responses associated with read versions of the utterances. This differentiation between the two versions holds up for the A condition as well as the UA condition although the difference is somewhat less when the utterances have been altered. CF had almost the same percent read scores for both versions in the UA condition, with a nonsignificant difference between the two versions for the A condition. If listeners respond in a dissimilar fashion to the two versions, they must be differentiating between them to some extent.

A more direct assessment of S/R identification ability can be made by considering the +, - and 0 CJDS based on percent correct scores. Examination of the percentage of

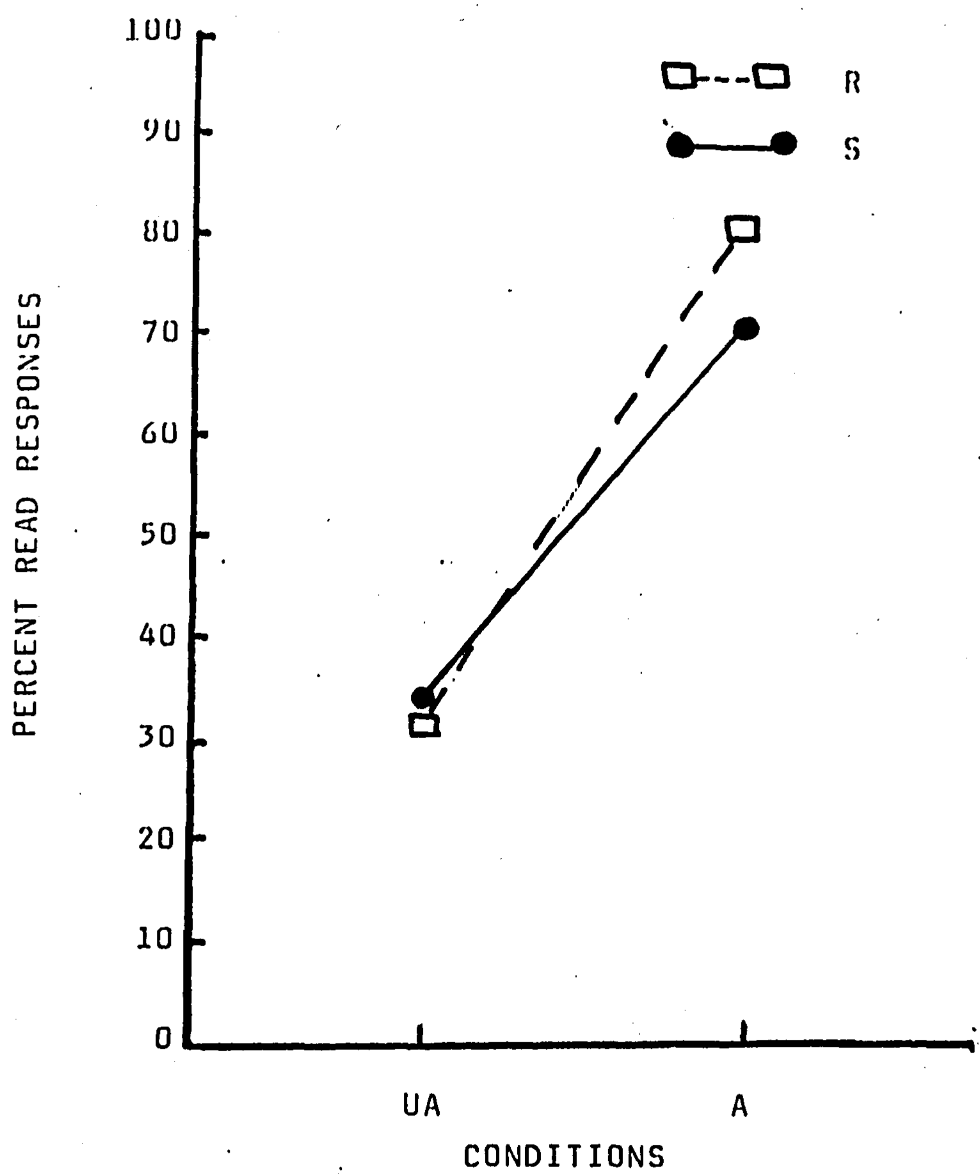


Figure 7. Percent read responses plotted as a function of unaltered and altered conditions for spontaneous and read versions for CF

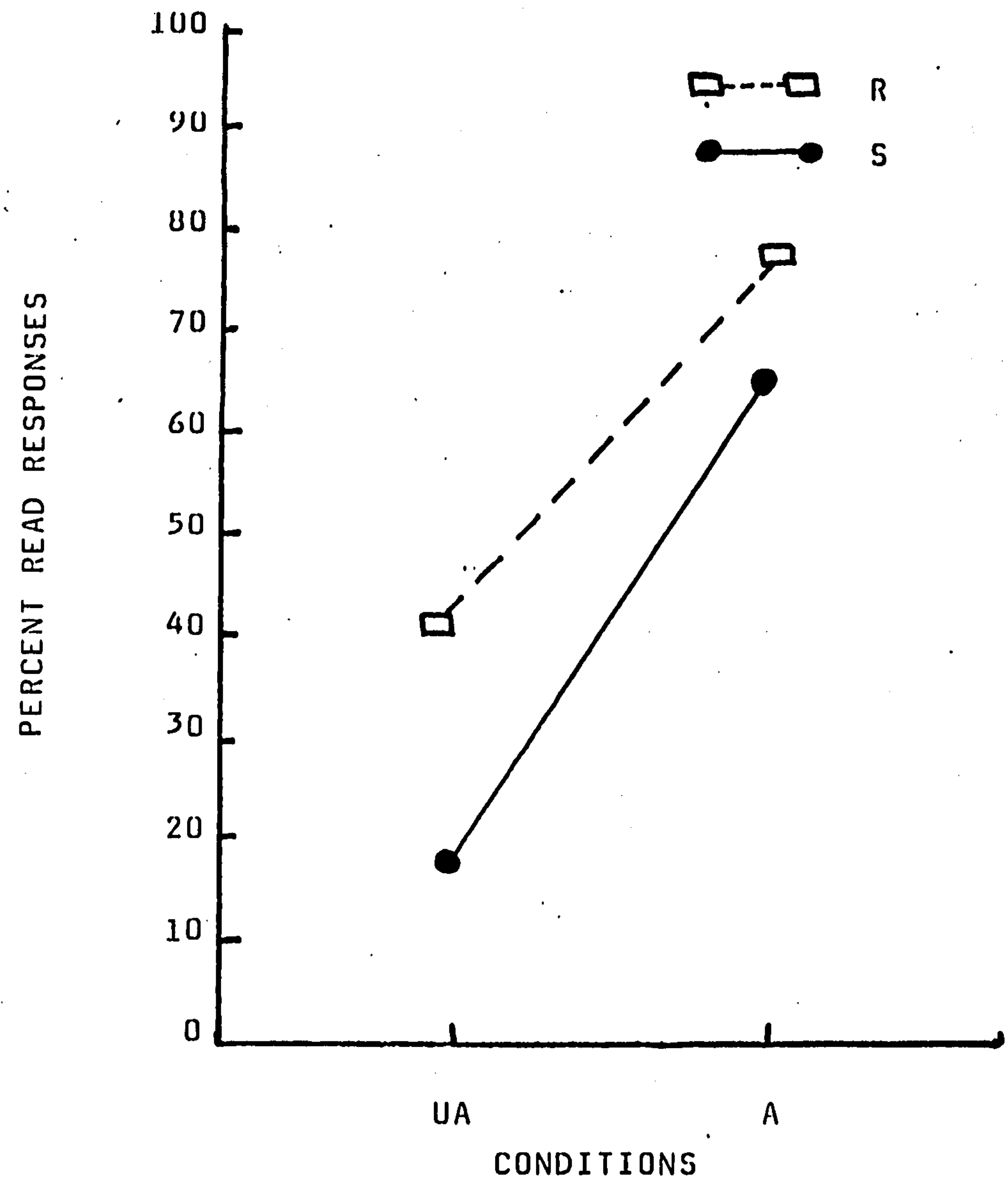


Figure 8. Percent read responses plotted as a function of unaltered and altered conditions for spontaneous and read versions for KM

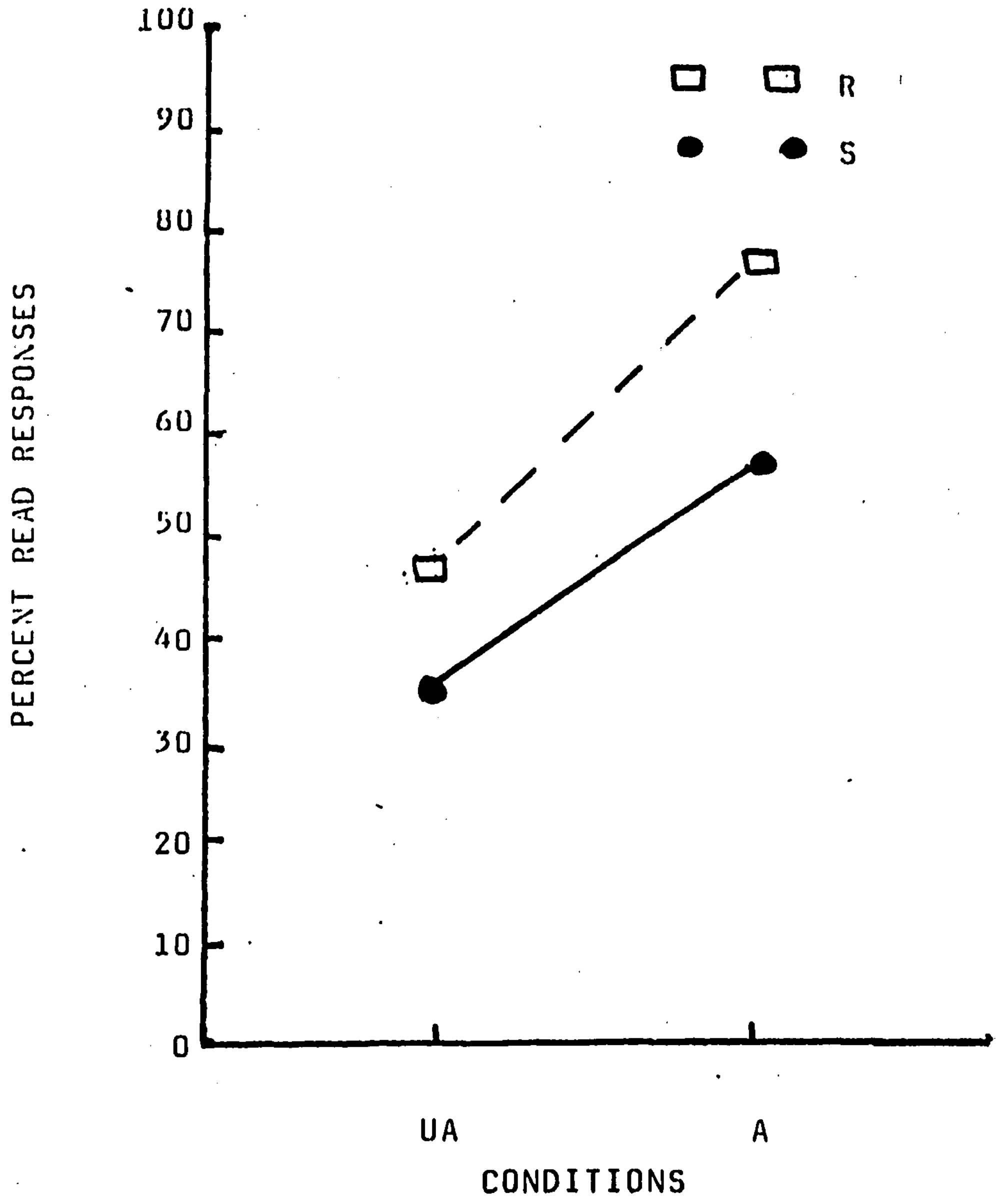


Figure 9. Percent read responses plotted as a function of unaltered and altered conditions for spontaneous and read versions for JP

unaltered versions which received the + designation in Table 8, revealed 53% for KM, 56% for JP and 23% for CF. Averaged across the three speakers, 44% of the utterances were identifiable. The low percentage of discriminable utterances perhaps reflects the fact that every utterance in connected discourse does not contain cues to the production mode. When listeners perceive speakers to be reading or speaking spontaneously, their perceptions are presumably based on salient cues within a few utterances.

The use of the conversational mode, in an attempt to encourage spontaneity, may have contributed to a limitation of the acoustic and perceptual separation of the two versions. Having reading material in the form of a conversation, particularly one originally produced by the speakers themselves, may have reduced the "readness" of the read material. Given the read material in script form may have encouraged the speakers to try, consciously or unconsciously, to sound spontaneous. The conversational mode was probably successful in achieving a high degree of spontaneity for the spontaneous versions. It may have inadvertently done the same for the read ones.

If a discrimination rather than an identification task had been used, a higher percentage of utterances might have received responses at greater than chance level. However, the single presentation method more clearly reflects what occurs in real life. A follow-up study to this investigation will include a perceptual task where the listeners will make discrimination judgments using the same utterances.

Listener ability to make S/R distinctions based on single version presentation raises an interesting question. How are listeners able to make this identification? References in the literature to variation between spontaneous and read utterances have always been to relative differences. One or the other of the two modes has been characterized as being faster, or higher pitched, or "more expressive." Implicit in these kinds of comments is that a comparison must be made in order to identify the speaking mode. Although the percentages of discriminable utterances were lower than anticipated, listeners were still able to make some correct decisions. Something other than a comparison analysis method must be invoked to account for these correct responses. Listeners apparently have an internal representation of spontaneity or "readness" that parallels that of phonemic and suprasegmental aspects of speech.

F_0 Contribution to S/R Identification

While the number of identifiable utterances was limited, the basis upon which these successful identifications was made is still of interest. Presumably the cues which allowed listeners to identify individual utterances are the same ones which operate in greater than utterance length speech samples. There are two kinds of evidence that point to the F_0 as primary locus of the S/R difference: (1) effects of contour flattening and (2) listener comments made after completion of the S/R discrimination task.

Results of F_0 Contour Flattening

The presence of the significant differences between the percent read responses for the two conditions, UA and A, suggests that listeners employed an F_0 based strategy in making S/R judgments. They identified flattened utterances as read significantly more often than they identified their unaltered counterparts as having been read (see Figures 1, 2 and 3). It can be assumed that in the absence of clues to the contrary (or at least salient ones) listeners tended to identify unaltered versions of the utterances as having been spontaneously produced. Altered, F_0 flattened utterances were more frequently identified as having been read. Such a strategy has two significant implications for the current investigation. First it implies that the distinction between these spontaneous and read utterances lies in the F_0 contour. Manipulation of the F_0 contour was sufficient to alter listener response patterns.

The second implication of the significant difference between conditions is that the S/R distinction is cued by degree of movement of the F_0 contour. A flattened F_0 contour can be considered a special and extreme case of decreased F_0 movement. There is thus an inverse relationship between percent of read responses and degree of movement of the F_0 contour. This F_0 based listener strategy leads to the conclusion that the animation, liveliness and expressiveness associated with spontaneous speech is directly related to movement of the F_0 contour.

In addition to analyzing the effects of contour flattening in terms of percent correct responses, the CJDs were also considered. CJDs were assigned to both the unaltered and altered versions of the utterances which were then grouped on the basis of the relationship between the CJDs for the two conditions. Of the nine possible combinations, the eight groups that emerged were: =0, ==, -0, --, -+, 00, 0+, 0-. There were no =- utterances. The first designation for each group is that for the UA condition; the second is the CJD for the A condition. The numbers and percentages of utterances in each of the CJD groups are given in Table 21.

What does each of these groupings suggest about the effect of F_0 flattening? There were seventeen utterances in the +0 group (29%) which were judged correctly at a greater than chance level under the UA condition, but only at chance level under the A condition. The implication is that when the utterances were unaltered listeners were able to identify them correctly on the basis of the F_0 contour. When this information was lost through contour flattening, listeners were no longer able to make these judgments correctly. The presence of utterances with this +0 pattern supports the hypothesis that the F_0 contour contains some of the information supporting the S/R distinction.

The four utterances in the -0 group (7%) also support the hypothesis that the F_0 contains the primary cue to S/R

TABLE 21.--Number of Utterances in each Chance Judgment Designation (CJD) Group by Speakers

CJD Group	Speaker CF	KM	JP	Total
+ 0	4	7	6	17
- 0	2	1	1	4
+ +	1	3	4	8
- -	1	1	0	2
- +	1	0	1	2
0 +	1	0	0	1
0 -	3	0	0	3
0 0	9	7	6	22
Total	22	19	18	59

differentiation. Although in the UA condition the listeners were making incorrect judgments, these judgments were made consistently (i.e., at greater than chance level) on the basis of presumed misinformation in the F_0 contour. When the contour was flattened listeners no longer judged correctly. Adding this 7% to the 29% in the +0 CJD group there is a total of 36% of the utterances that provide support for the hypothesis that S/R discrimination information is contained in the F_0 contour.

These data indicate that for somewhat more than one-third of all the utterances F_0 contour is important in the S/R dichotomy. Of the 33 utterances which were identified at greater than chance level in the UA condition the +0 and -0 groups combined make up 63%.

What do the other categories suggest about the role of F_0 in S/R identification? Seventeen percent of the utterances fell into either the ++ (14%) or the -- (3%) categories. They were identified at greater than chance level (either correctly or incorrectly) and the identification ability was unaffected by F_0 contour flattening. There are two possible explanations for utterances falling into the ++ and -- groups. These utterances could have been cued by other aspects of the speech signal than F_0. In that case flattening the contour would not affect discrimination. Another possibility is that F_0 cues coexist with durational or spectral ones which supported the distinction when F_0 information was eliminated.

There is a group of six utterances (10% of the corpus) which has different designations from one another, but which are classified together because of their anomalous nature. These are -+, 0+ and 0-. Inclusion in one of these three groups suggests that listeners either went from consistently incorrect judgments to consistently correct judgments with flattening of the F_0 contour, or that the consistency of judgments occurred after the F_0 contour was flattened. It is interesting to note that five of these six sentences were produced by speaker CF. The presence of most of the anomalous utterances among those of CF is further evidence that her utterances differ in some essential way from those of the other two speakers. The final anomalous utterance was one of JP's.

Listener Comments

After completion of the spontaneous/read identification task participants were asked to indicate the bases upon which their judgments were made. Listeners were given no indication of what appropriate criteria might be. They were simply asked to list the criteria they used. (See "Instructions for Spontaneous/Read Identification Task" in the Appendix.) The comments were grouped into criterion categories and the attribute associated with the two versions of the sentences noted. This information is included in Table 22 along with the number of times that a given criterion was stated. A number of listeners mentioned more than one category which accounts for there being a total of 85 responses for 40 listeners.

TABLE 22.--Criteria Indicated by Listeners as Bases for Perceptual Judgments in S/R Identification Task; Attributes associated with each Version; Number of Times each Criterion Mentioned and Percentage of Total Comments for each Criterion

Criterion	Spontaneous Attribute	Read Attribute	Number of times Mentioned	Percentage of Total[a]
Intonation, inflection	Greater inflect.	Monotonous	34	40
Stress pattern,emphasis	Greater emphasis	Less emphasis	21	25
Fluency; presence & placement of pauses	Stumbled before words; Fluidity; breath at non-punctuation points	Breath's only at commas	10	12
Pitch level	Elevated pitch	Lower pitch	5	6
Degree of expressive-ness; emotion	A lot more feeling; trace of emotion in voice	More mechanical	5	6
Timing; speaking rate	More natural timing	Speeded up	4	5
Presence o absence of stressed syllables	Stressed syllables	No stressed syl-lables	2	2
Sentence endings	Greater terminal pitch variation; complete sentence endings	Abrupt sentence endings	1	1
Presence of echo	No echo	Echo	1	1
How sounded when re-peated to self	-----	-----	1	1
Style	-----	-----	1	1

[a]There were only 40 listeners, but 85 comments because some listeners listed more than one criterion

These data show that 55 (65%) of the comments were "intonation", "inflection", "stress pattern", or "emphasis." Of these 55, 34 (40%) mentioned intonation or inflection specifically. Because intonation is so closely related to sentence stress patterns it can be assumed that references to stress and emphasis are to the same aspects of the speech signal as "intonation" and "inflection."

Five respondents indicated that "pitch level" was used as a criterion in making S/R decisions, higher pitch being associated with spontaneity. It was assumed that this variable referred to overall or average pitch level rather than to its contour.

Given that the listeners (a) heard only a single utterance at a time, and (b) only heard a single version of each utterance, no relative pitch information was available. How could they make judgments about pitch level? A single speaker indicated that she used the criterion of how the sentence sounded when she repeated it to herself. This listener may have been the only one to verbalize a process that others were using -- referring back to their own production mechanisms. The same kind of analysis-by-synthesis mechanism could be invoked to explain correct S/R judgments with a single version presentation mode.

There were only four instances (5%) where duration was mentioned. Of the remaining 29%, 12% were related to degree of fluency or the placement of pauses. There were only two utterances that contained pauses. These were CF8 and KM6. Listeners who mentioned this variable may have been referring to their strategy with these utterances or they may have been responding to pseudo-pauses or non-pauses perceived as pauses.

The criterion categories "degree of expressiveness" (6%), "presence or absence of stressed syllables" (2%), "sentence endings" (1%), and "style" (1%) might be related to the F_0 contour, although it is not clear that they are. The criterion category "presence of echo" probably referred to a perceptual "hollowness" that characterized the F_0 contour flattened utterances. These listener comments strongly implicate F_0 contour in the S/R distinction.

These two pieces of evidence--the effect of F_0 flattening and the listener comments--support the hypothesis that listeners use F_0 and its contour to make S/R judgments. The significance of these data must be considered cautiously. Because of the nature of the task, one of judging utterances whose F_0 contours had been manipulated, it is possible that F_0 may have assumed increased salience. Since the listener comments were elicited after a forty minute exposure to altered versus unaltered utterances, listeners may have become artificially sensitized to F_0. Duration and spectral factors that varied between versions might subsequently have been ignored due to a spuriously salient F_0. With hindsight it is obvious that the question of S/R identification criteria might reasonably have been posed both before and after the listening task.

A similar kind of artificiality may have been operational when listeners were making S/R judgments. Listeners commented that they had a great deal of difficulty making decisions. Since they were forced to choose, they may have grabbed onto the most convenient handle--the state of the F_0 contour--but one that was contrary to what might be used under ordinary circumstances. Although increased salience may have been a factor in the results, both subjective and objective data point very strongly to the primacy of F_0 in S/R identification.

Status of Acoustic Differences Between Versions

It is reasonably clear from the data discussed above that some of the differences between spontaneous and read versions of the utterances reside in the F_0 contour. The data on the acoustic differences, however, fail to isolate the specific variable or variables responsible. Certain significant differences appeared when F_0 measurements were made, but not in the same variables in the same direction for all three speakers. Average F_0 was higher a significant number of times for one version than the other for all three speakers. However, the direction differed. For JP and KM spontaneous utterances were more frequently higher than read ones. CF's read utterances were more frequently higher than her spontaneous ones.

The movement of the topline appeared to be a significant variable for speaker JP. This was reflected both in slope direction measurements and significantly higher DRRs for his read utterances. For KM, PRR appears to be a significant variable in the S/R distinction, with greater movement of the peaks for his spontaneous utterances. It appears then, that no one of the variables measured consistently separated spontaneous and read utterances for each of the speakers.

Duration measurements did not differ consistently between versions. The lack of consistent variability between versions in overall duration is surprising as there are numerous references in the literature to speaking rate differences as a function of formality level. Changes in speaking rate should be reflected in overall duration. There are several possible reasons why there were not consistent overall duration differences between versions for the present corpus. The first possible reason is that noise obscured differences that actually existed. Given the wide range of syntactic variation among the utterances, and the small number of speakers used, this might easily have been the case. A second possible explanation for there being no overall duration differences is that perception of rate variation among levels of formality is a function of the number and length of pauses rather than of changes in duration. Confounding of duration and pauses is inherent in speaking rate measures of the number of linguistic segments per unit time. A third possibility is that variation in speaking rate is an idiosyncratic variable which was not functional for the three speakers used in this study. Whatever the explanation, neither overall duration, the FSR, nor the SUR varied consistently between versions.

Thirty-seven percent of the utterances had what might be considered atypical contours, that is their toplines did not decline. Although there are numerous references in the literature to F_0 contour variation as a function of differences in focus and syntactic structure, the contours almost always have falling slopes. Because of the intimate relationship between the syntactic structure of an utterance and its F_0 contour, an attempt was made to assess whether syntactic structure might account for the atypical F_0 contours. (A structural breakdown of the utterances appears in the Appendix.) This was done by identifying the types of phrases (e.g. adjective phrase, prepositional phrase, etc.) into which each utterance could be broken down. The kinds of phrases associated with typical contour utterances were compared with those associated with atypical contour utterances. There was no obvious difference between the phrase types of the group of utterances having typical contours and those utterances having atypical contours.

The number of non-declining toplines for this corpus can probably be accounted for on the basis of the conversational context from which the utterances were taken. The F_0 contours presumably reflected contextual references, speaker intent, relationship between utterances and speaker intention. These contours are departures from "neutral intonation contours" associated with reading (usually of isolated read sentences or monologue type narratives) where no particular emphasis is indicated. Under more natural speaking conditions non-declining F_0 contours might be more representative than neutral intonation contours.

A further attempt to tease out the specific F_0 variable(s) used in making S/R judgments was made through calculating the correlation between "percent read" judgments and each of the quantifiable F_0 variables: average F_0, F_0 range, the DRR and the PRR. Table 23 contains the correlation coefficients for each of the variables by speakers and versions. As was the case with the acoustic measurements themselves, no clear picture developed. All of the coefficients were extremely small, indicative at best of weak correlations.

On the basis of results for the individual speakers negative correlations would be expected for average F_0, F_0 range and the PRR. That is these variables would be expected to decrease as the percent read judgments increased. A positive relationship would be anticipated for the DRR since greater declination was associated with "readness." For F_0 average a single weak correlation occurred for the read utterances of JP (-.42). There were three weak negative correlations for F_0 range: for the read utterances of CF (-.49), and for both spontaneous and read versions for KM

TABLE 23.--Coefficients for Correlations between Percent Read Responses and Each of the Four Quantifiable F_0 Variables by Speakers and Versions

	Speakers					
	CF		KM		JP	
Versions	S	R	S	R	S	R
F_0 Variable						
Average F_0	-0.08	-0.05	-0.22	-0.07	0.32	-0.42
F_0 Range	-0.09	-0.49	-0.38	-0.38	0.12	-0.04
Declination Range Ratio	0.37	0.00	0.13	0.06	0.20	0.31
Peak Range Ratio	0.33	-0.14	0.15	0.39	0.21	0.29

(both -.38). Obviously the correlations are too low to support strong conclusions.

Dispersion of Data Points

There was a strong tendency for a wider dispersion of data points to occur with the acoustic measurement data of the spontaneous versions as compared to those of their read counterparts. This was the case for 15 of the 21 measurements made across speakers. Of the remaining six measurements, three had nearly equal dispersions for the measurements of the two versions. Only three had wider dispersions for the read measurements over those for the spontaneous measurements.

The nature of this greater dispersion for the spontaneous data was the existence of a few extreme data points, rather than a spreading out over the whole range. Most of the data points for both spontaneous and read versions clustered about an average point. This dissimilarity in dispersion suggests that one of the differences between spontaneous and read speech is that during the former speaking mode various acoustic parameters display a greater tendency to deviate from the average values of that parameter. In a relatively long speech sample this tendency might not be realized for all the utterances. Only some of them would contain the cues for spontaneity. This marking of only a portion of the utterances as spontaneous is consistent with the finding that only a portion of the utterances were identifiable.

Speaker and Listener Idiosyncracy

At the onset of this study the concept of spontaneity and "readness" was considered to be relatively simple and straightforward. It was assumed that these two speaking modes represented different points along a formality continuum. Their placement would be speaker and situation dependent, but predictable. All other things being equal, it was assumed that spontaneous speech would be less formal than read speech and that acoustic aspects of the two modes would differ in predictable ways. The use of lexically identical utterances in the present experiment was an attempt to keep all other things equal. However, the resulting picture of spontaneous and read differences did not approximate what was anticipated.

Three very different pictures were presented by each speaker, suggesting that the formality/informality relationship between spontaneous and read speech is not straightforward. JP's spontaneous speech sounded very casual; his read speech seemed prototypically read. KM's speech samples were very different from those of JP. Although the whole read conversation had a more "read" quality than the spontaneous one, individual excised utterances did not sound as readily identifiable as did those of JP. All of the version pairs sounded different, that is they would be discriminable on a discrimination task. It was surprising, however, that many of them were identifiable as to speaking mode.

In contrast to JP and KM, CF's utterances were anomalous. Her spontaneous utterances seemed less casual than her read ones. This difference may have been an artifact associated with the recording sequence. The spontaneous sample was elicited during her initial visit to the location. She may have been more relaxed, and hence less formal, when she read the transcribed conversation. If the differences between spontaneous and read speech were robust, however, this should not have mattered. These observations of speaker idiosyncracy suggest that different speakers handle the S/R distinction in different ways. A formality/informality distinction is obviously not isomorphic with the S/R continuum.

In addition to speaker and situational effects on the S/R distinction, a listener variable may also exist. Although there was no analysis of listener responses, listener comments suggest the use of different response protocols. Forty percent of the listeners' comments were related to intonation or inflection. However, fluency, stress pattern, pitch level, timing and sentence endings were also mentioned. Different listeners, who mark "readness" in their own speech in various ways, may listen to different aspects of the speech signal when making S/R judgments. When Nakatani and Aston (1981) investigated factors affecting lexical stress perception, they noted that certain listeners seemed to attend to pitch differences while others listened for duration differences. A similar listener variable may have been at work in the S/R perceptual task.

Although no note was made of individual listener responses in degree of certainty, the casual observation was made while scoring that some listeners seemed more certain of their responses than other listeners did. This may have been a reflection of a personality variable related to self-confidence or some other psychological trait. However, it also might have reflected dissimilarities in listening criteria. Those listeners who were less certain of their responses might have been so because they did not hear the kinds of cues they were listening for, the ones that have the greatest salience for them.

In conclusion it appears that defining utterances in terms of perceived spontaneity is much more complicated than first conceived. Rather than a one dimensional S/R continuum which lines up in some way with formality/informality, it is likely that utterances are located in a multidimensional space by reference to sets of coordinates that include idiosyncratic speaker variations, listener dissimilarities, sex of the speaker, and a host of other unidentified variables.

Implications for Speech Synthesis

As synthesis of natural sounding speech becomes more successful the need will arise for synthetic speech to reflect nuances associated with extra-linguistic factors. One such factor may be whether the speech was read or produced spontaneously. Up to this point, the meager references in the literature have suggested that spontaneous and read speech differ primarily in nature of pauses, the degree to which certain phonetic and phonological processes occur, and in speaking rate.

In attempting to synthesize speech which will give an impression of spontaneity, it may then be necessary to alter the F_0 contour in some way. Perhaps the topline should decline less to give the impression of greater spontaneity. If a conversational mode is to be portrayed, it may be necessary to include utterances with topline flatness or rises (atypical contours). Synthesis of speech in a conversation mode may sound more natural and convincing if all the contours are not declining. In order to indicate interaction among speakers, manipulation of the F_0 contour away from the neutral form may be necessary. Relationships between and among communicants in conversations may be indicated by subtle differences in the F_0 contour. There also may have to be greater movement of the F_0 peaks over the F_0 range.

Summary and Conclusions

The answer to the question, "Can listeners correctly identify isolated utterances as having been read or spoken spontaneously?" is "Yes, sometimes." For the present corpus, 42% of the utterances across speakers were identifiable according to an established criterion. What this figure suggests is that not all utterances within a connected narrative will be marked for speaking mode. Rather, a certain proportion of them might be so marked, with listeners basing their S/R judgments on this subset of utterances.

In order to produce synthetic speech which is identifiable as having been produced through a given speaking mode, it will probably be necessary to mark certain utterances with the acoustic lineaments peculiar to that mode. What are the acoustic features that mark speech as spontaneous or read? It seems clear that the S/R distinction is cued primarily by F_0 and its contour. Listener strategy, listener comments and the effects of contour flattening support the F_0 cue hypothesis. Duration factors measured in this study appeared not to play a role in the distinction.

Although the role of F_0 in the S/R distinction for the present corpus is clear, the specific aspects of F_0 which vary and the ways in which they vary are obscure. Average F_0 surfaced as the single parameter that differed significantly between versions for all three speakers, but the direction of the differences varied among speakers. The

PRR for KM and the DRR for JP differed significantly between versions. This suggests that either the range of peak movement or the movement of the topline may cue the S/R distinction. Some of JP's versions appeared to be distinguishable with reference to the slope. The slope of his utterances more frequently changed between versions than did those of the other speakers, and declination appeared to be more clearly represented for his read versions. Idiosyncracy appears to be the prevailing characteristic associated with these data.

In the conversational narrative type speaking mode used for the present experiment the neutral declarative F_0 contour was frequently inappropriate. Due to semantic highlighting and signalling of continuation, many of the F_0 contours' toplines either did not show the declination effect or they actually had rising toplines.

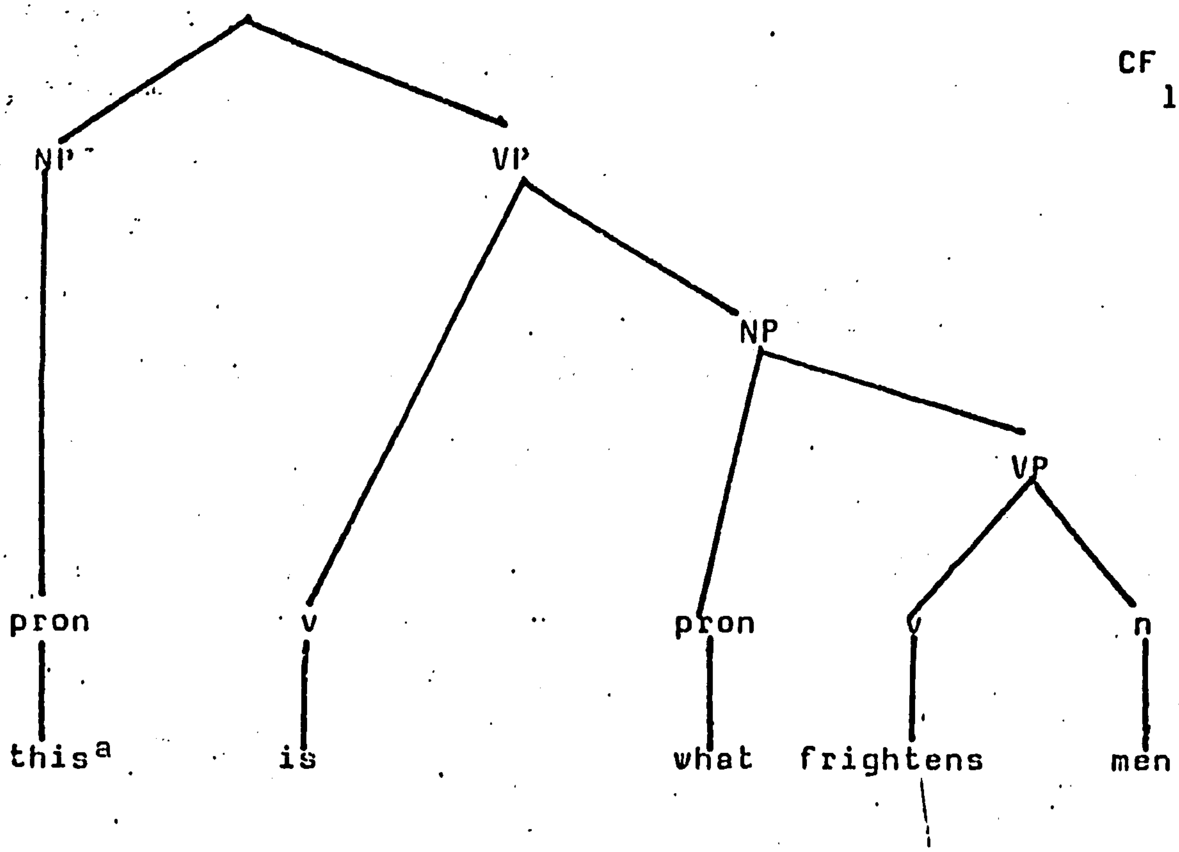
The present investigation was seen as an initial step in development of a model of differences between speech which is read and that which is produced spontaneously. From such a model could be generated an algorithm for manipulating perceived spontaneity or "readness." Some information leading toward this long range goal has been obtained through this investigation. The next step will be to isolate large numbers of utterances which are marked for the S/R distinction for comparison with those which are unmarked. Through this comparison it should be possible to catalogue the idiosyncratic features which cause listeners to decide whether an utterance has been read or whether it has been produced spontaneously.

APPENDICES

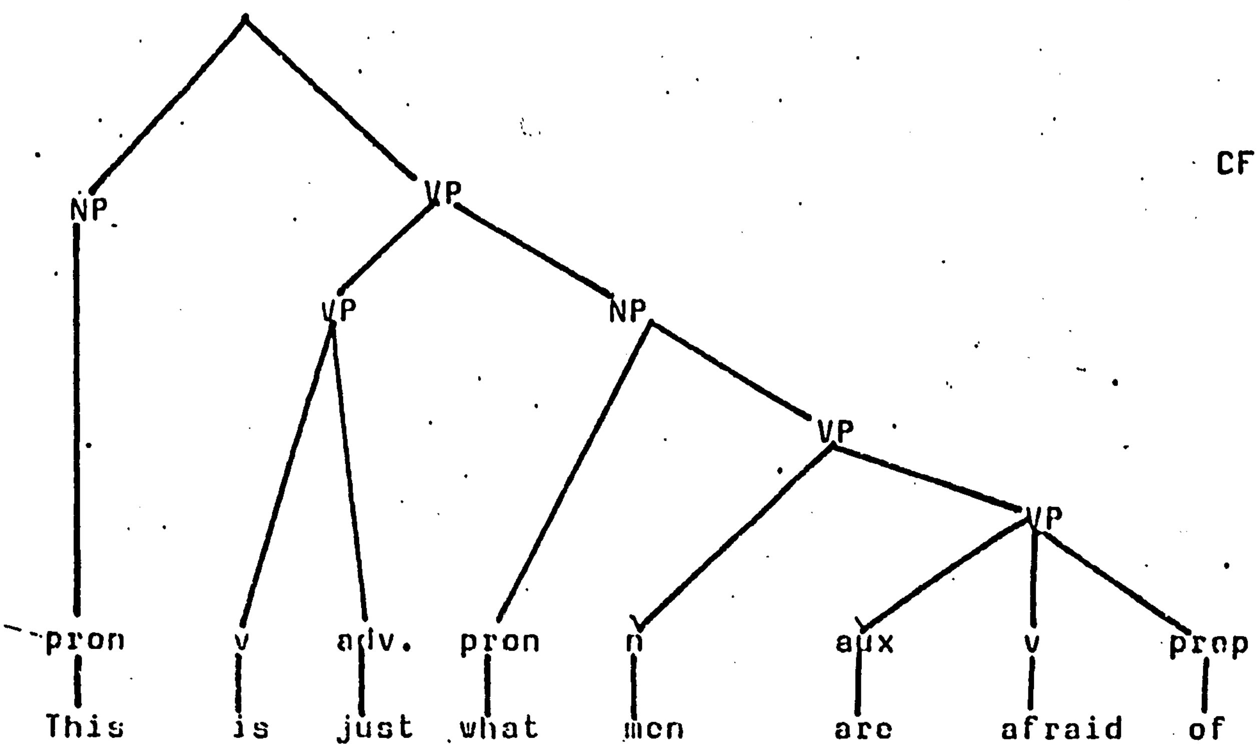
APPENDIX A

STRUCTURAL BREAKDOWN OF COMPLETE SENTENCE UTTERANCES

CF 1

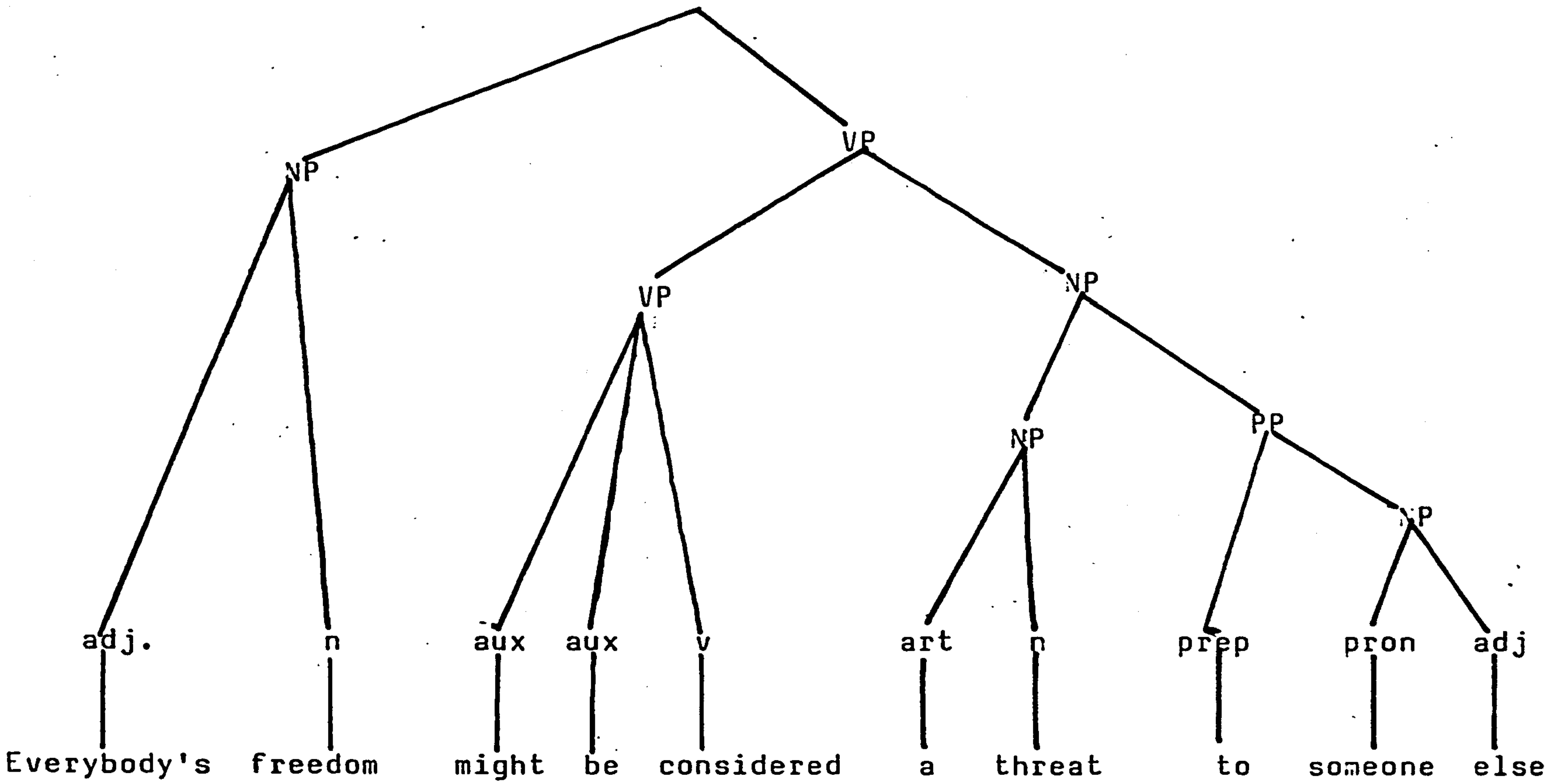


CF 3



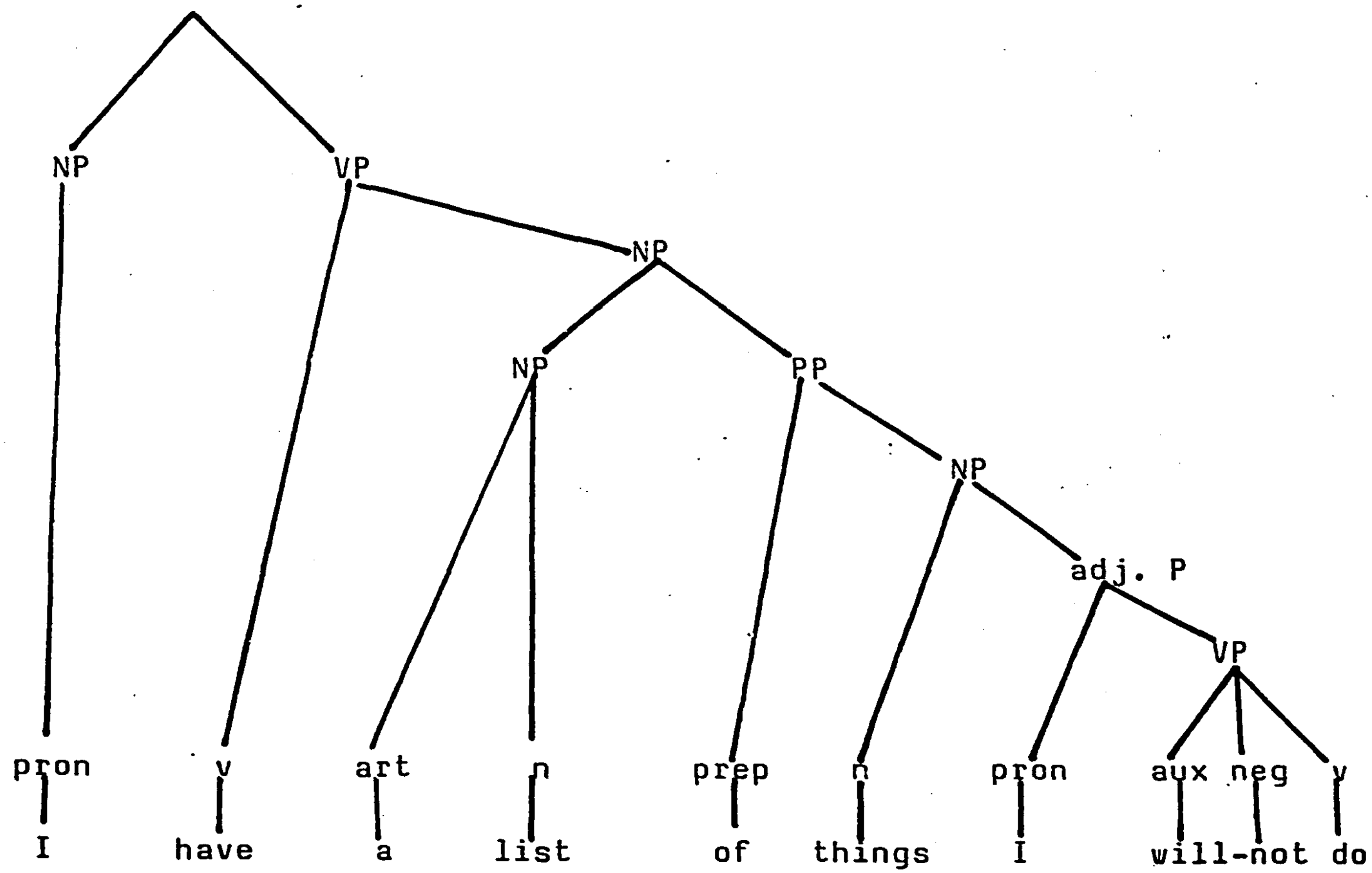
[a] Stressed syllables; ( ) indicates stressed in read version.

CF_4



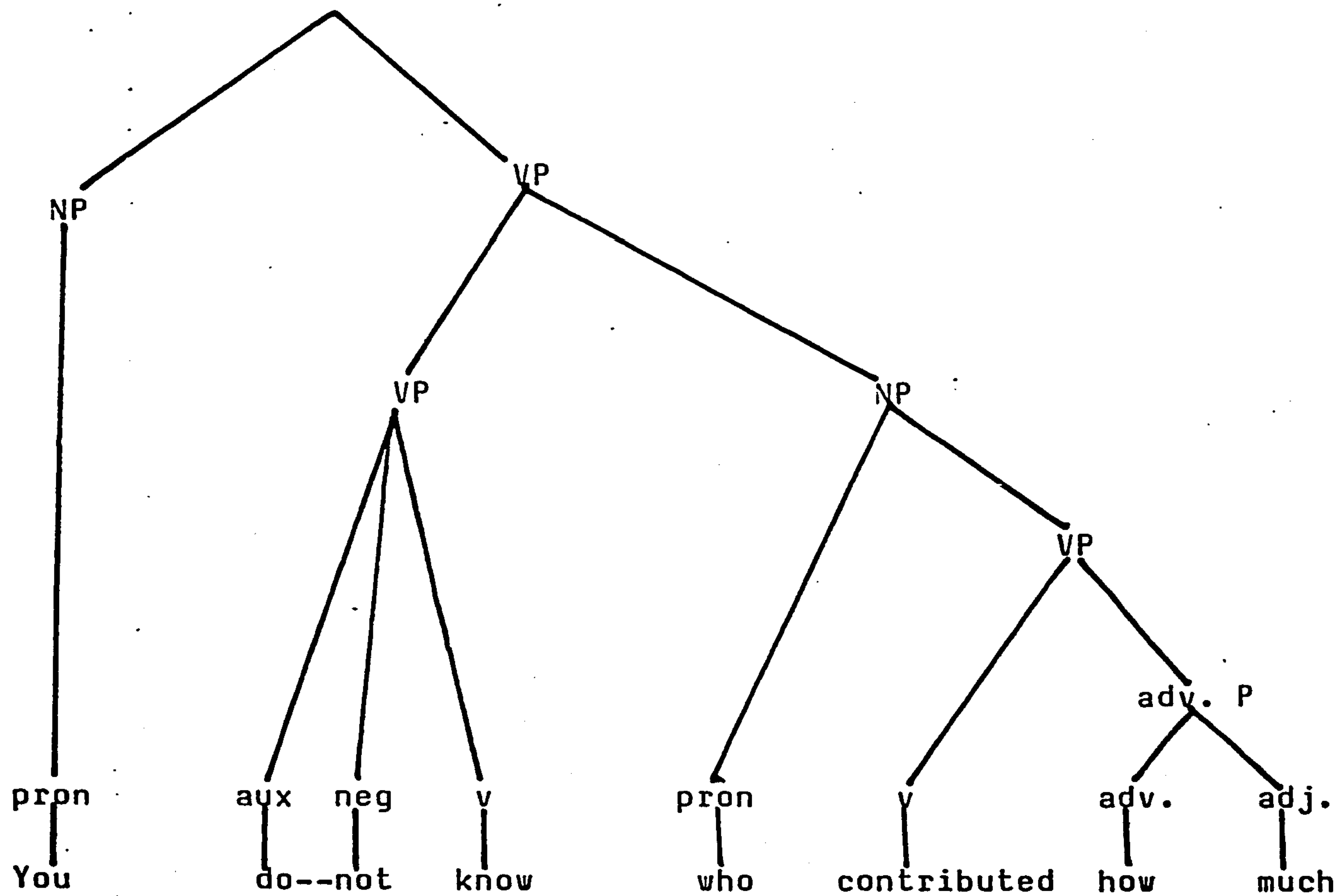
VP
NP
VP
NP
NP
PP
NP
adj.
n
aux
aux
v
art
n
prep
pron
adj
Everybody's
freedom
might
be
considered
a
threat
to
someone
else

CF 7

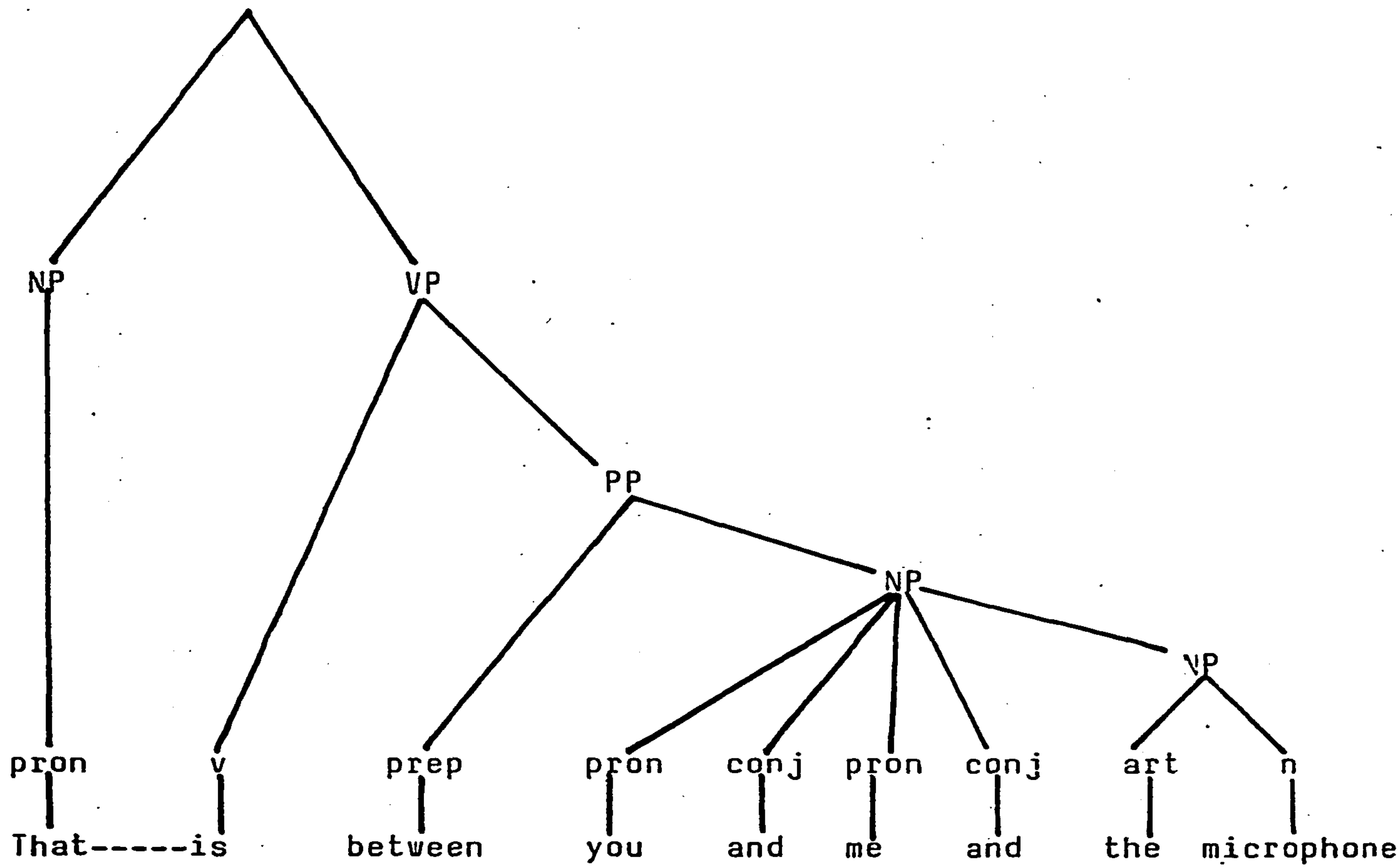


NP
VP
NP
NP
PP
NP
adj. P
VP
pron
v
art
n
prep
n
pron
aux
neg
v
I
have
a
list
of
things
I
will-not
do

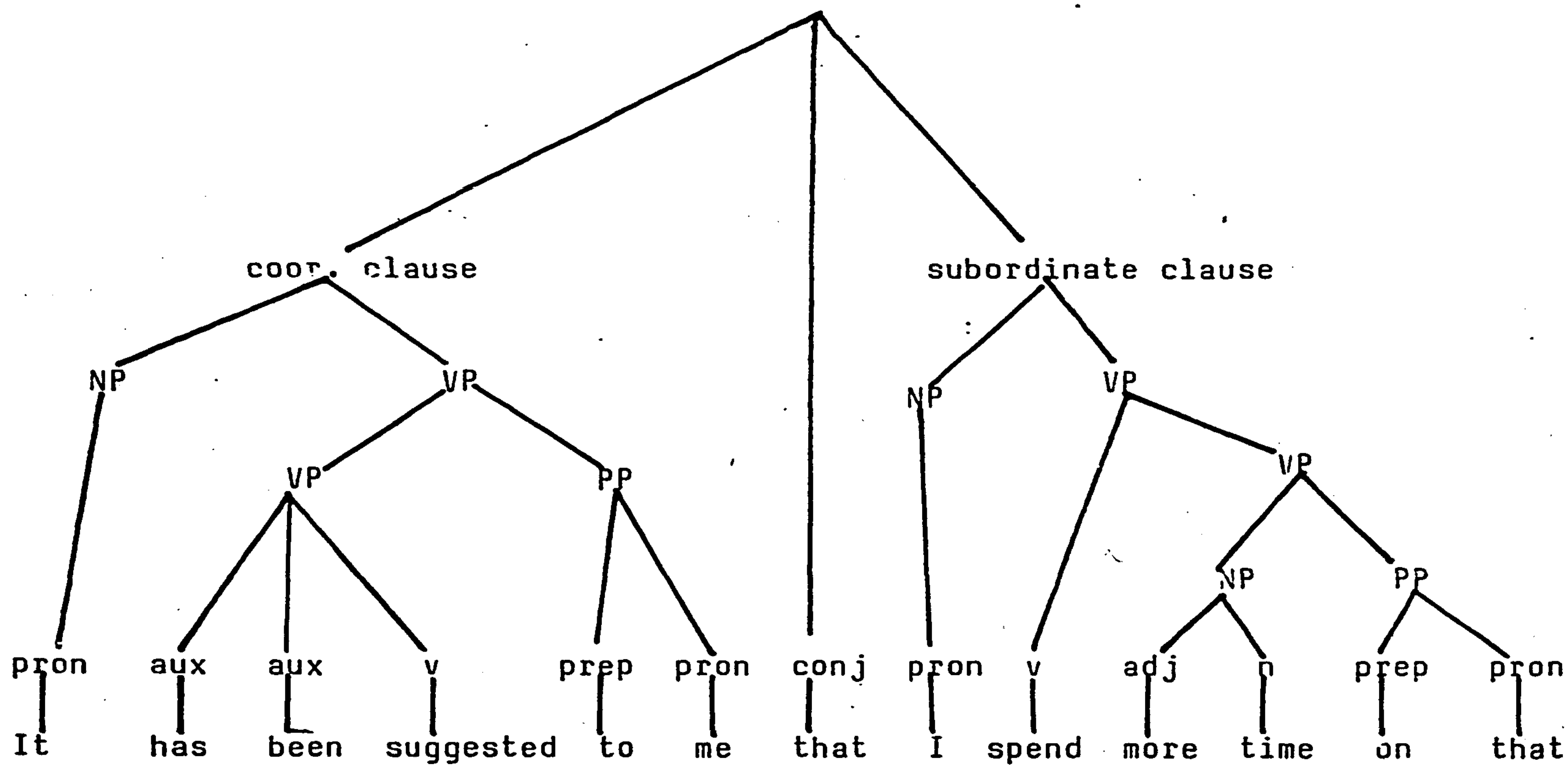
CF_8


NP
VP
VP
NP
VP
adv. P
pron
aux
neg
v
pron
v
adv.
adj.
You
do--not
know
who
contributed
how
much

CF_9

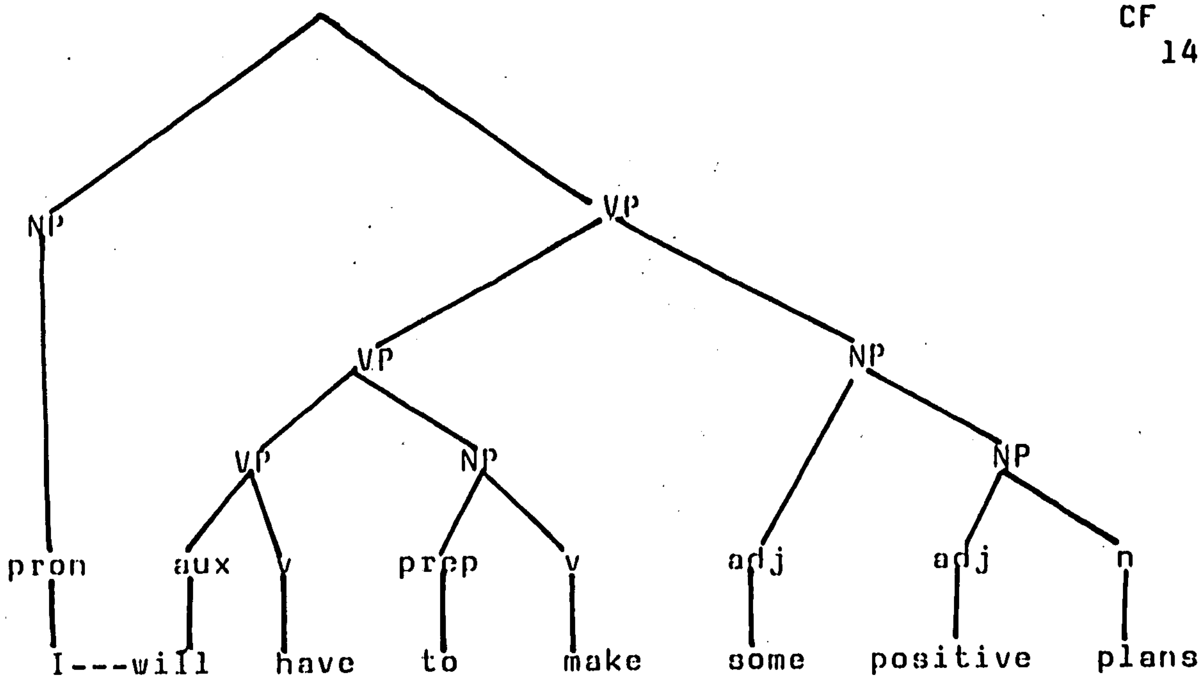

NP
VP
PP
NP
VP
pron
v
prep
pron
conj
pron
conj
art
n
That-----is
between
you
and
me
and
the
microphone

CF
13

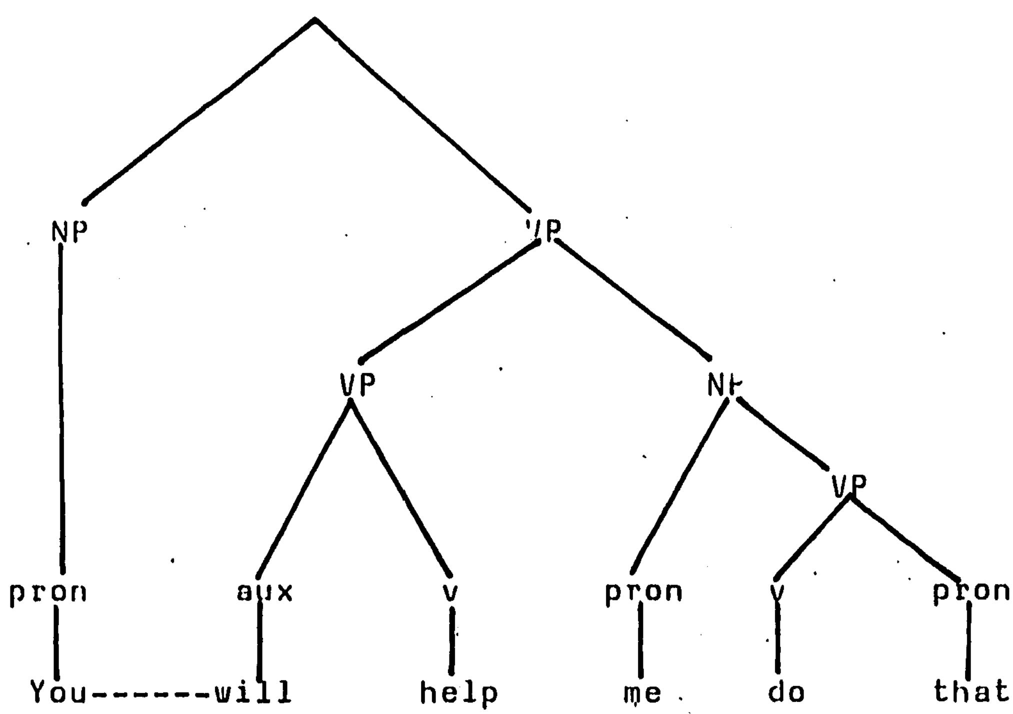


coor. clause
subordinate clause
NP
VP
NP
VP
VP
PP
VP
NP
PP
pron
aux
aux
v
prep
pron
conj
pron
v
adj
n
prep
pron
It has been suggested to me that I spend more time on that

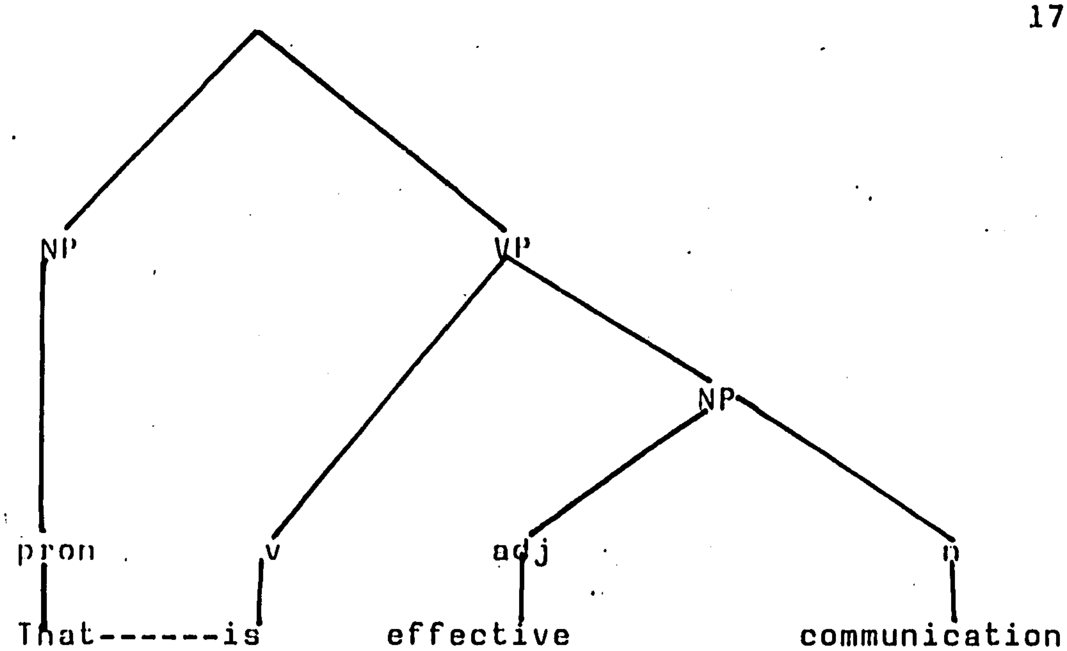
CF 14


NP
VP
VP
NP
VP
NP
NP
pron
aux
v
prep
v
adj
adj
n
I---will
have
to
make
some
positive
plans

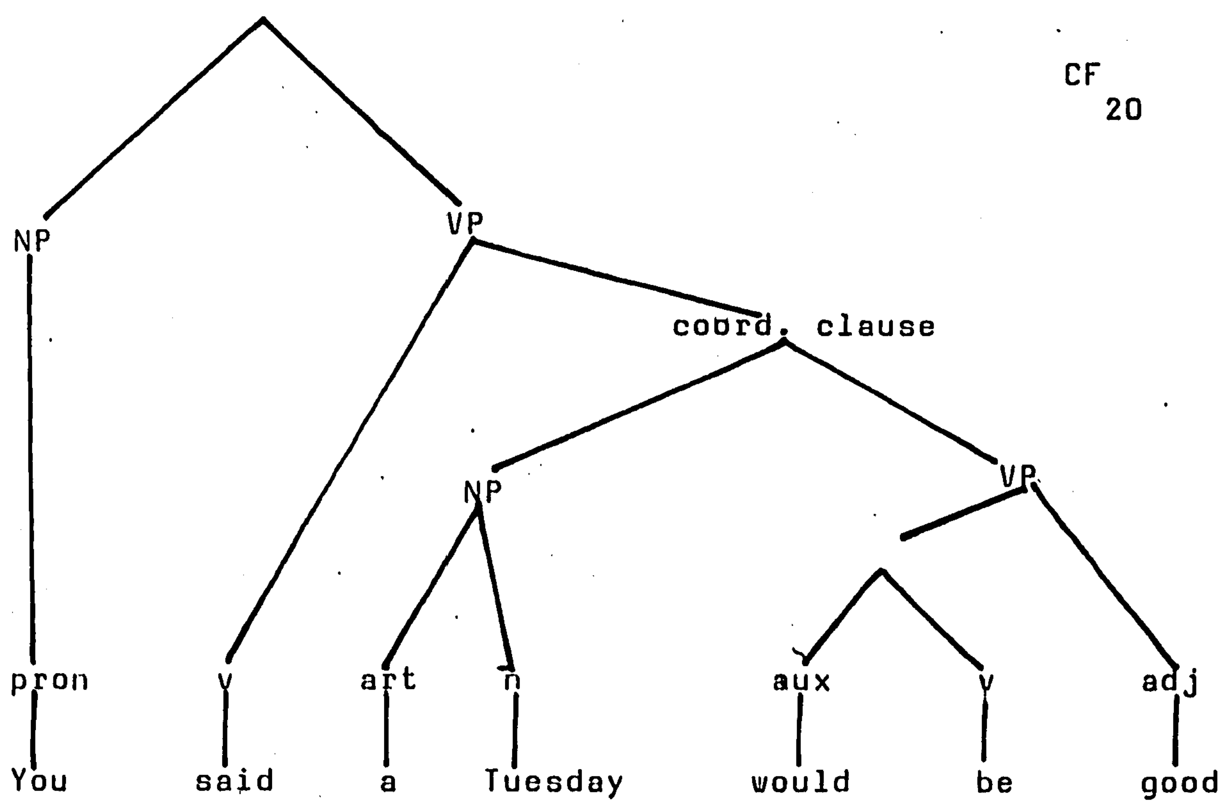
CF 16


NP
VP
VP
NP
VP
pron
aux
v
pron
v
pron
You------will
help
me
do
that

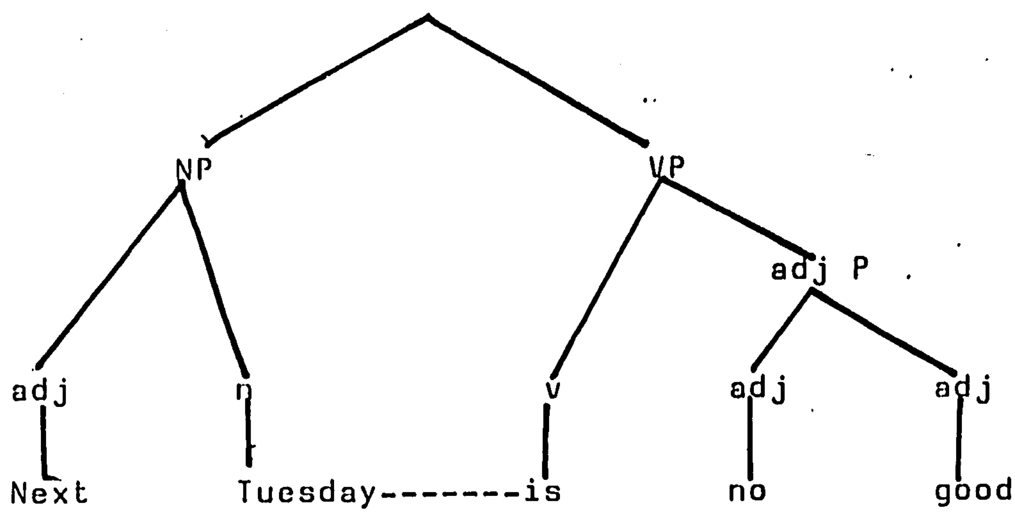
CF 17


NP
VP
NP
pron
v
adj
n
That------is
effective
communication

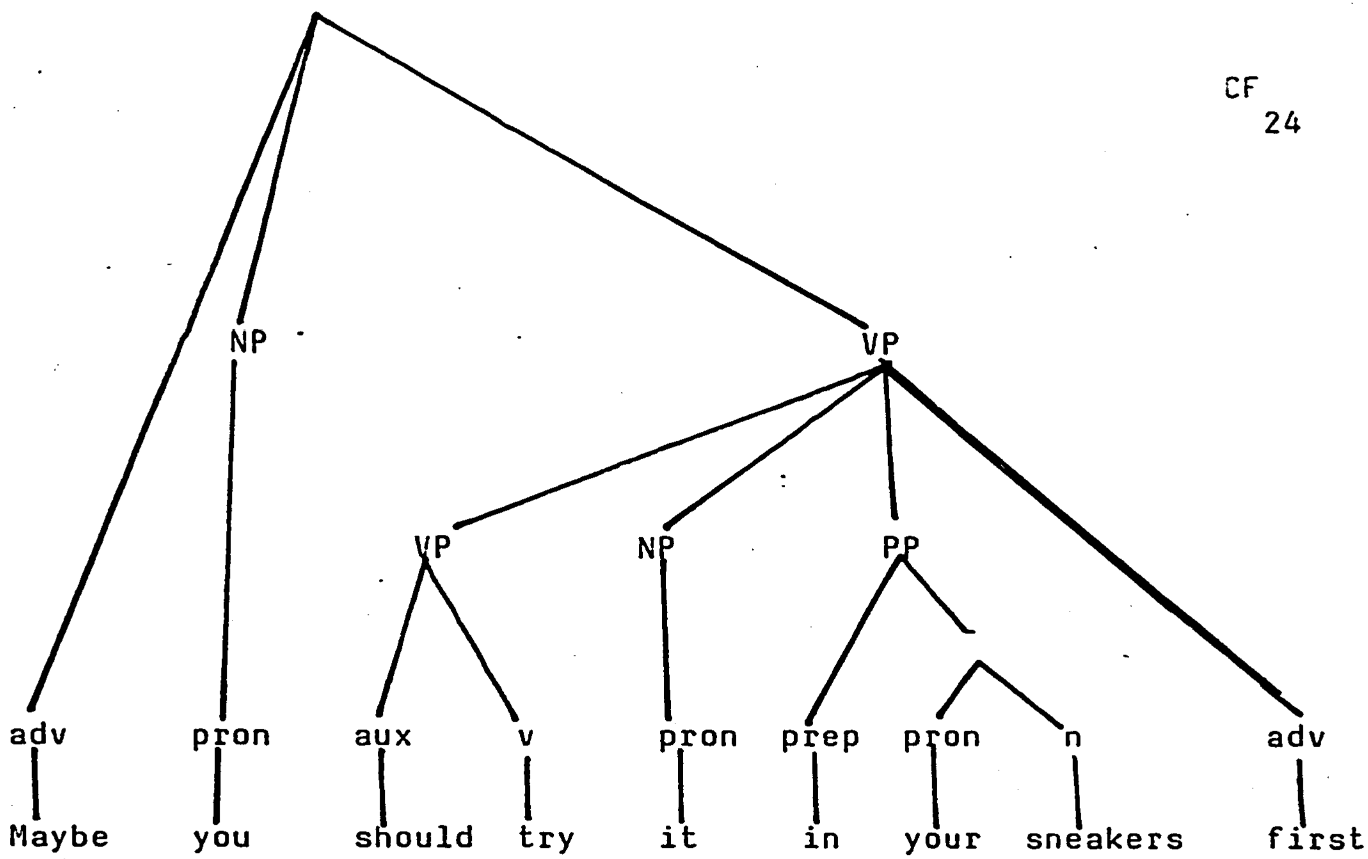
CF 20


NP
VP
coord. clause
NP
VP
pron
v
art
n
aux
v
adj
You
said
a
Tuesday
would
be
good

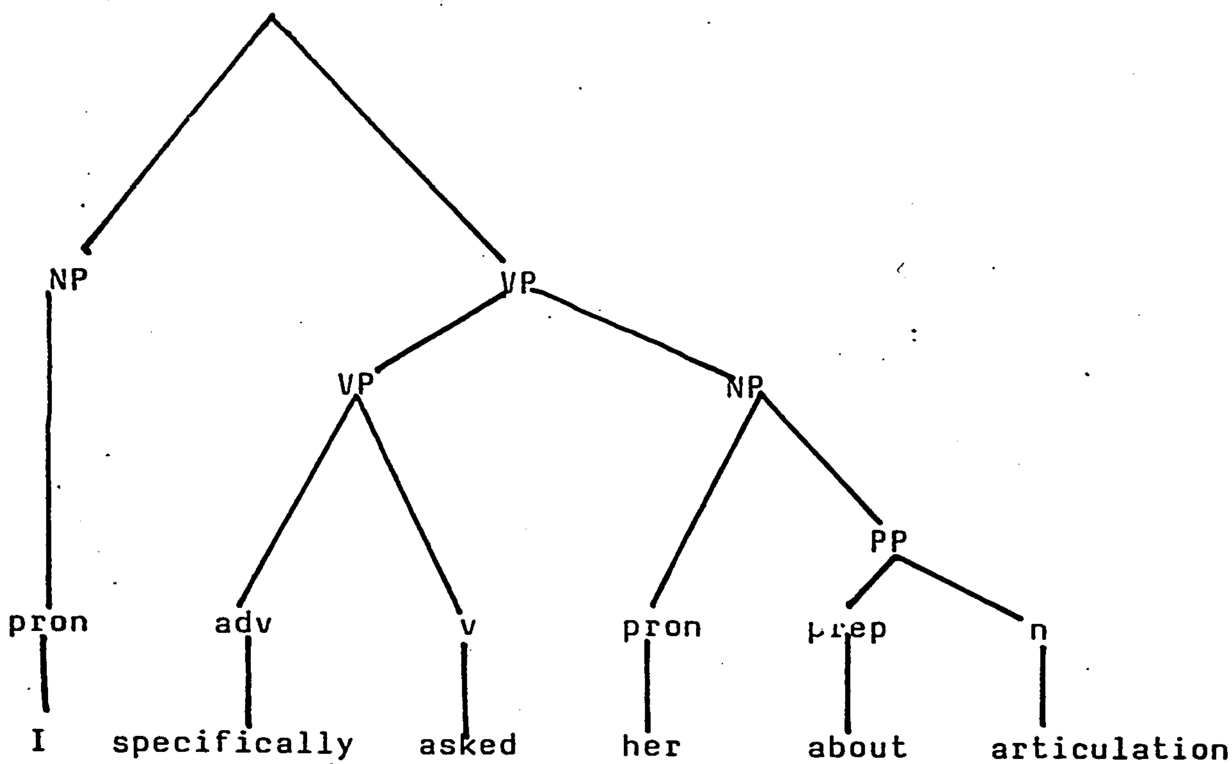
CF_{21}



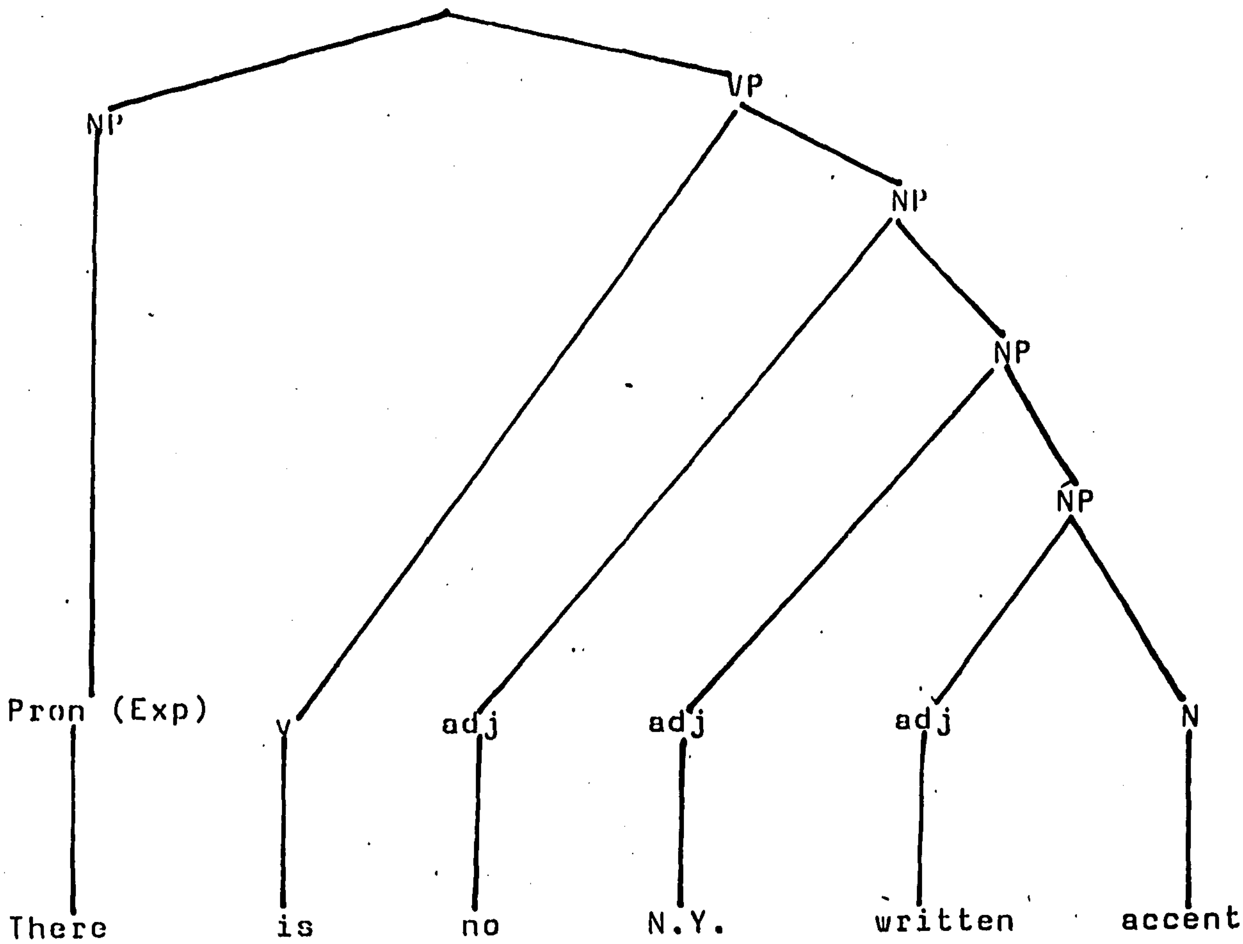
CF 24


NP
VP
VP
NP
PP
adv
pron
aux
v
pron
prep
pron
n
adv
Maybe
you
should
try
it
in
your
sneakers
first

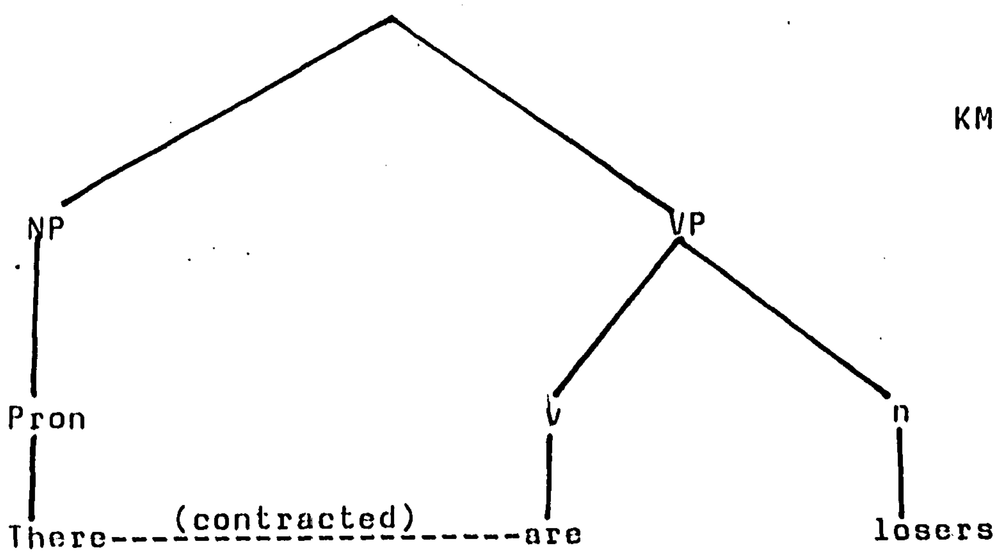
CF_{25}


NP
VP
VP
NP
PP
pron
adv
v
pron
prep
n
I
specifically
asked
her
about
articulation

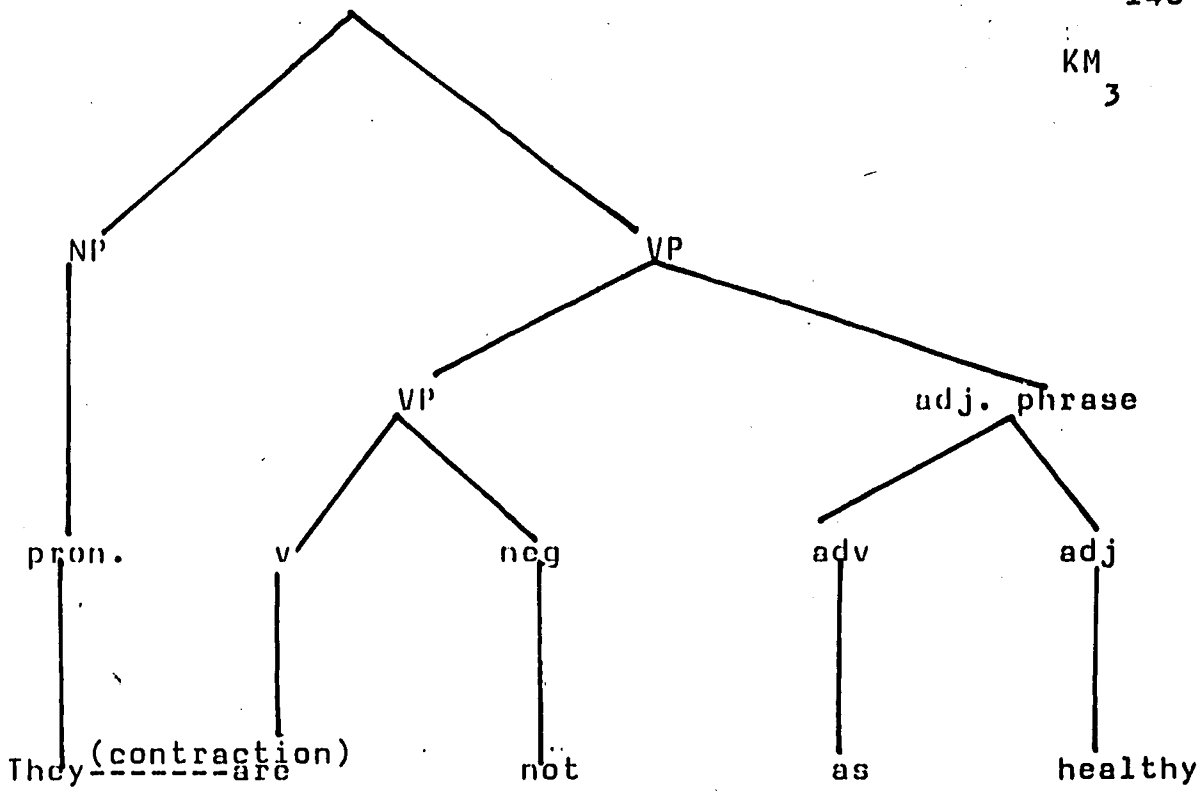
KM 1


VP
NP
NP
NP
NP
Pron (Exp)
V
adj
adj
adj
N
There
is
no
N.Y.
written
accent

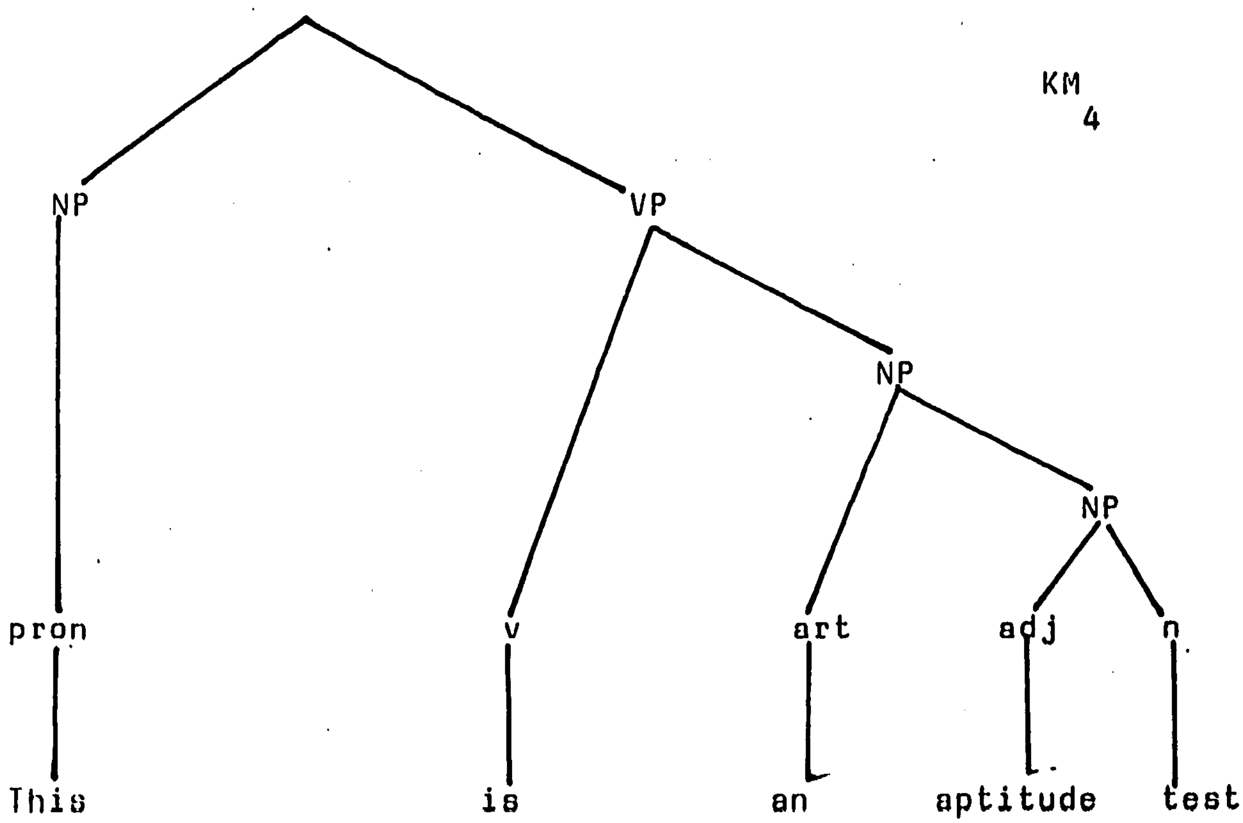
KM 2


NP
VP
Pron
V
n
There
(contracted)
are
losers

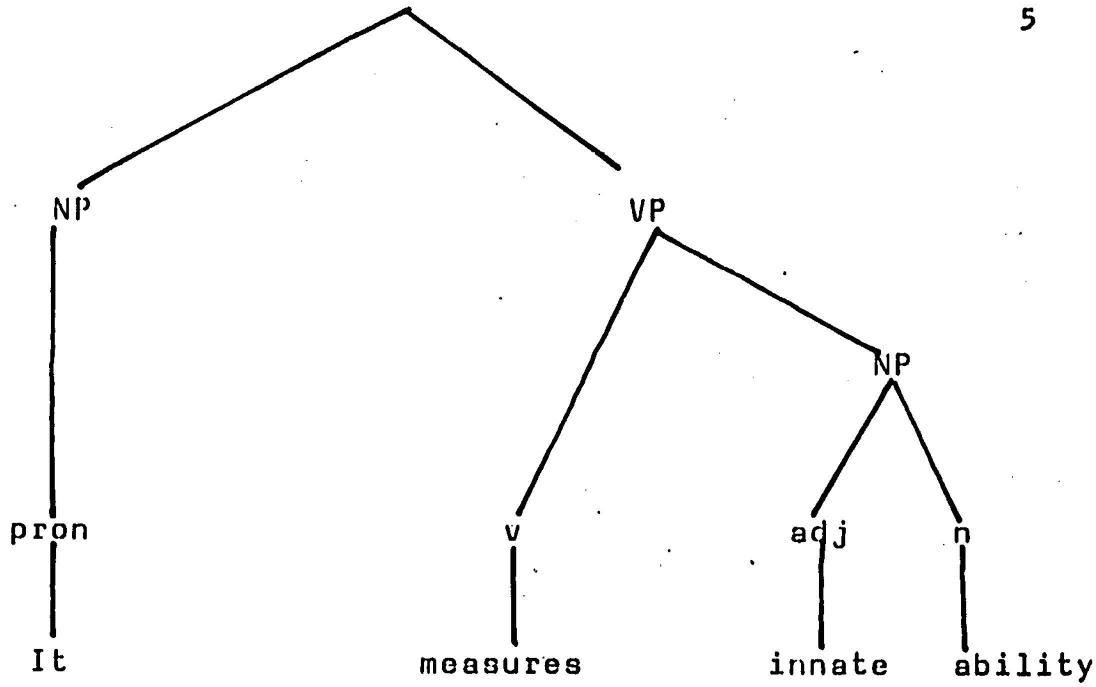
KM 3


NP
VP
VP
adj. phrase
pron.
v
neg
adv
adj
They
(contraction)
are
not
as
healthy

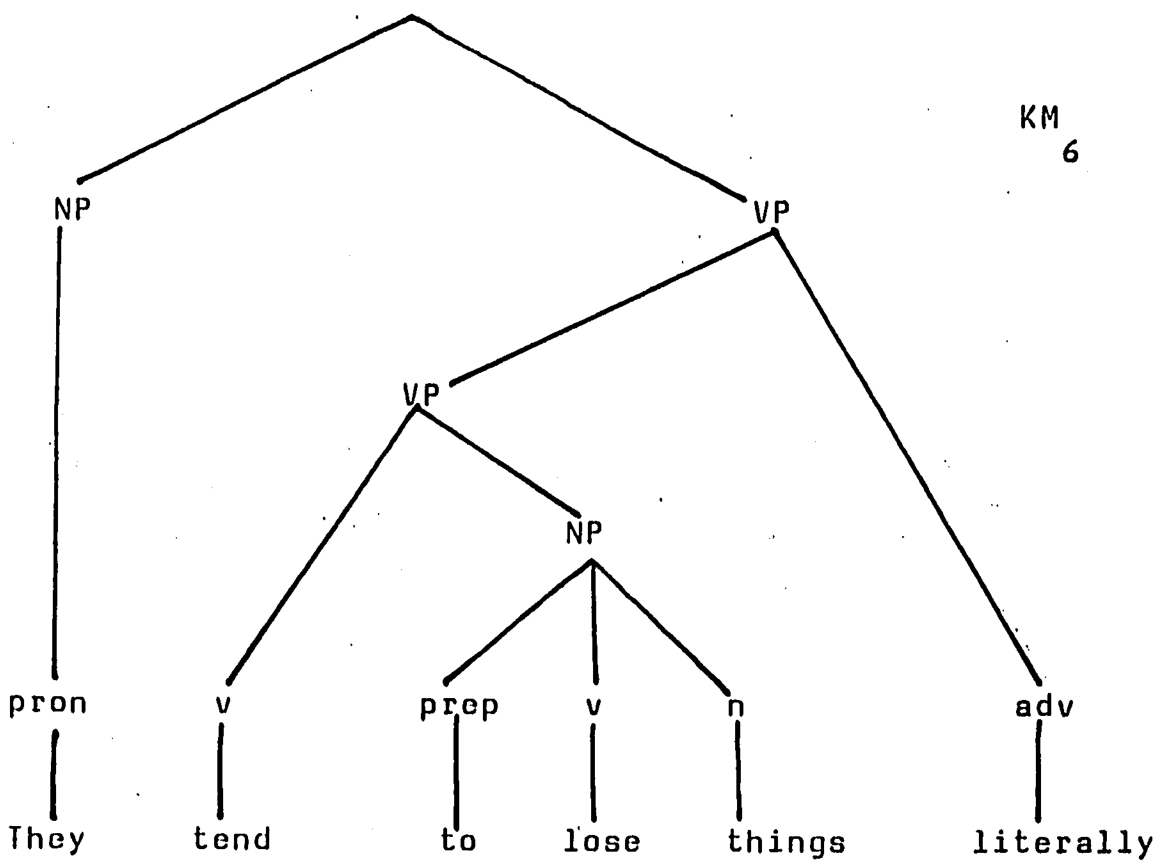
KM 4


NP
VP
NP
NP
pron
v
art
adj
n
This
is
an
aptitude
test

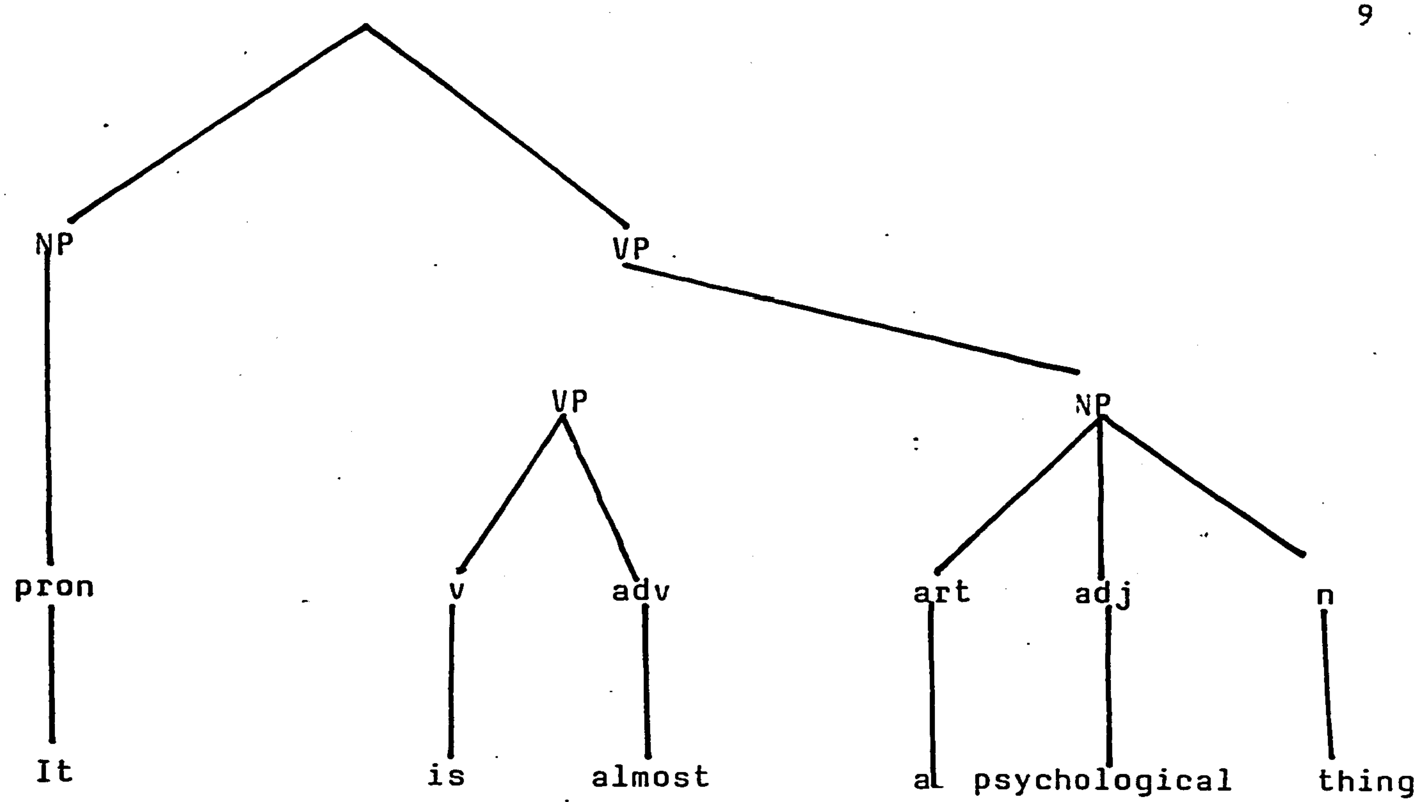
KM 5


NP
VP
NP
pron
v
adj
n
It
measures
innate
ability

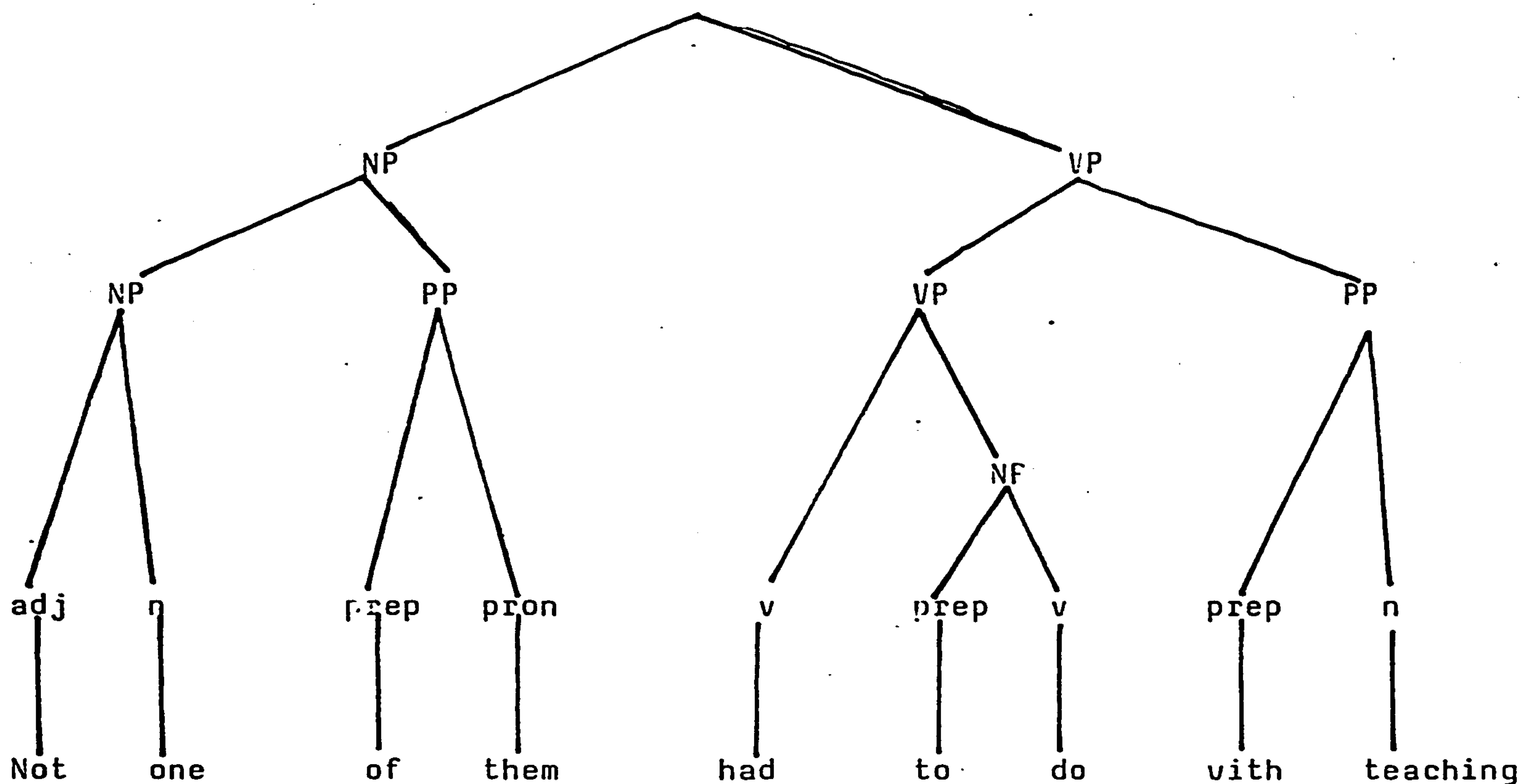
KM 6


NP
VP
VP
NP
pron
v
prep
v
n
adv
They
tend
to
lose
things
literally

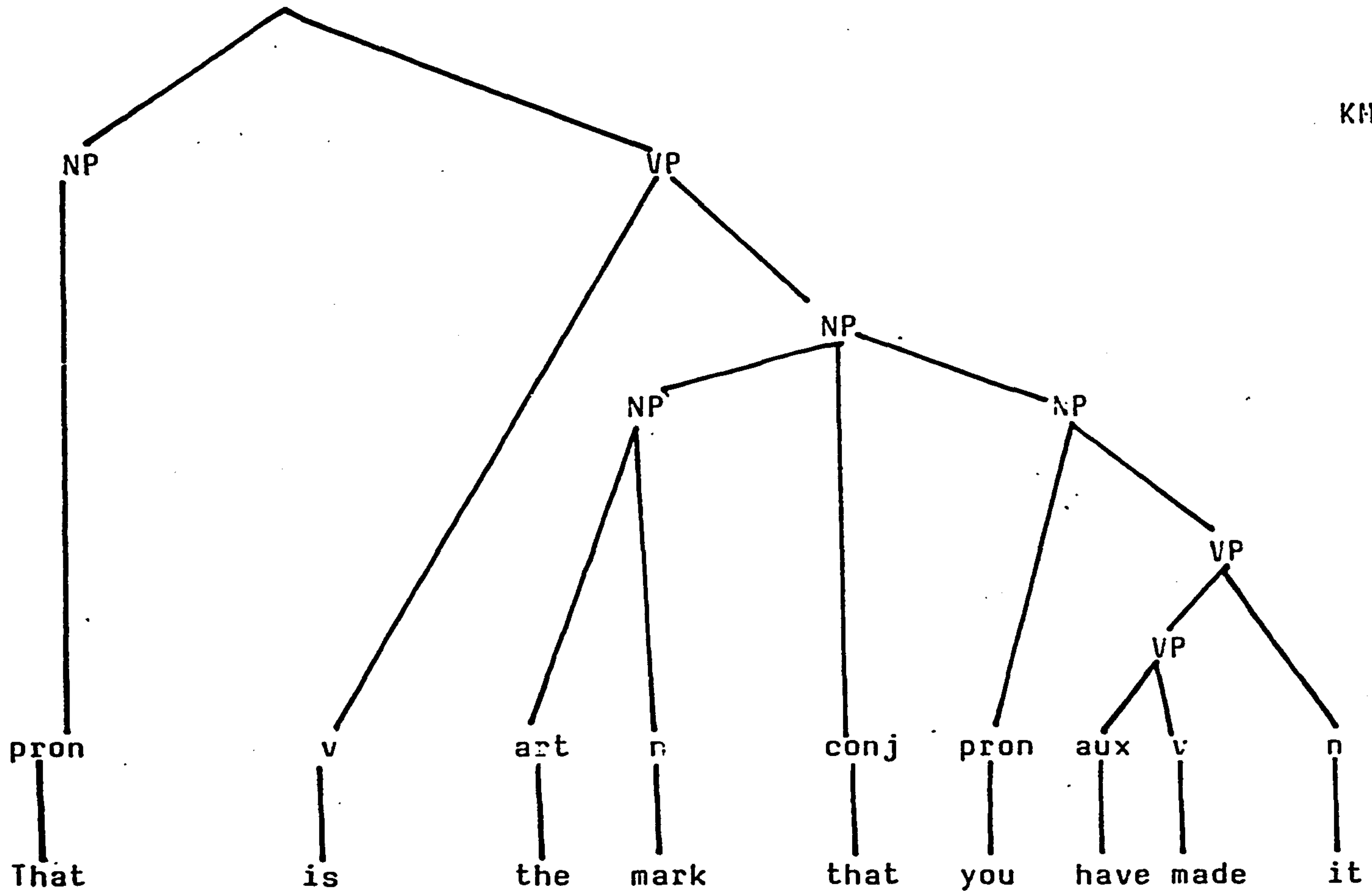
KM
9


NP
VP
VP
NP
pron
v
adv
art
adj
n
It
is
almost
a
psychological
thing

KM
10

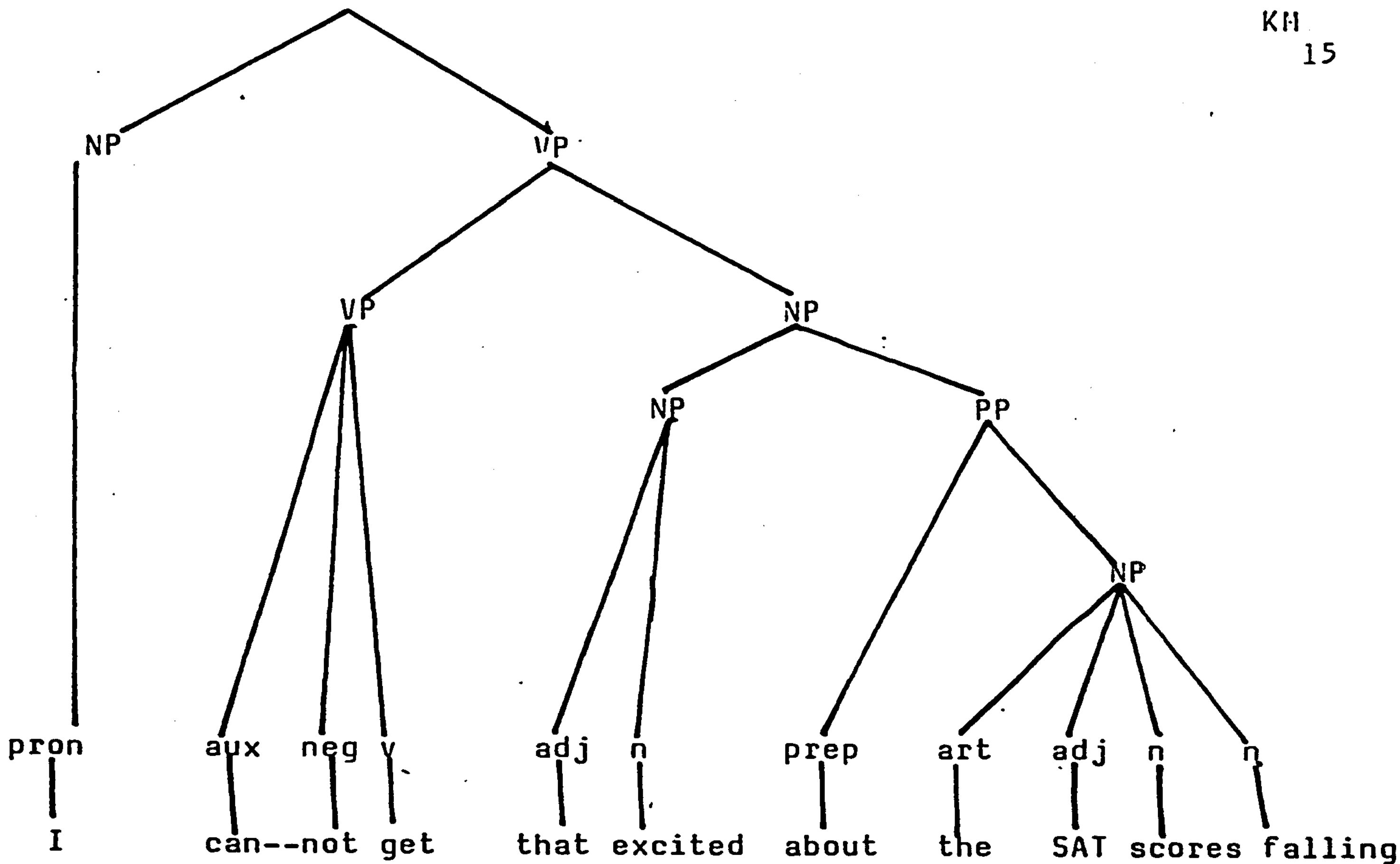


NP
VP
NP
PP
VP
PP
NP
adj
n
prep
pron
v
prep
v
prep
n
Not
one
of
them
had
to
do
with
teaching

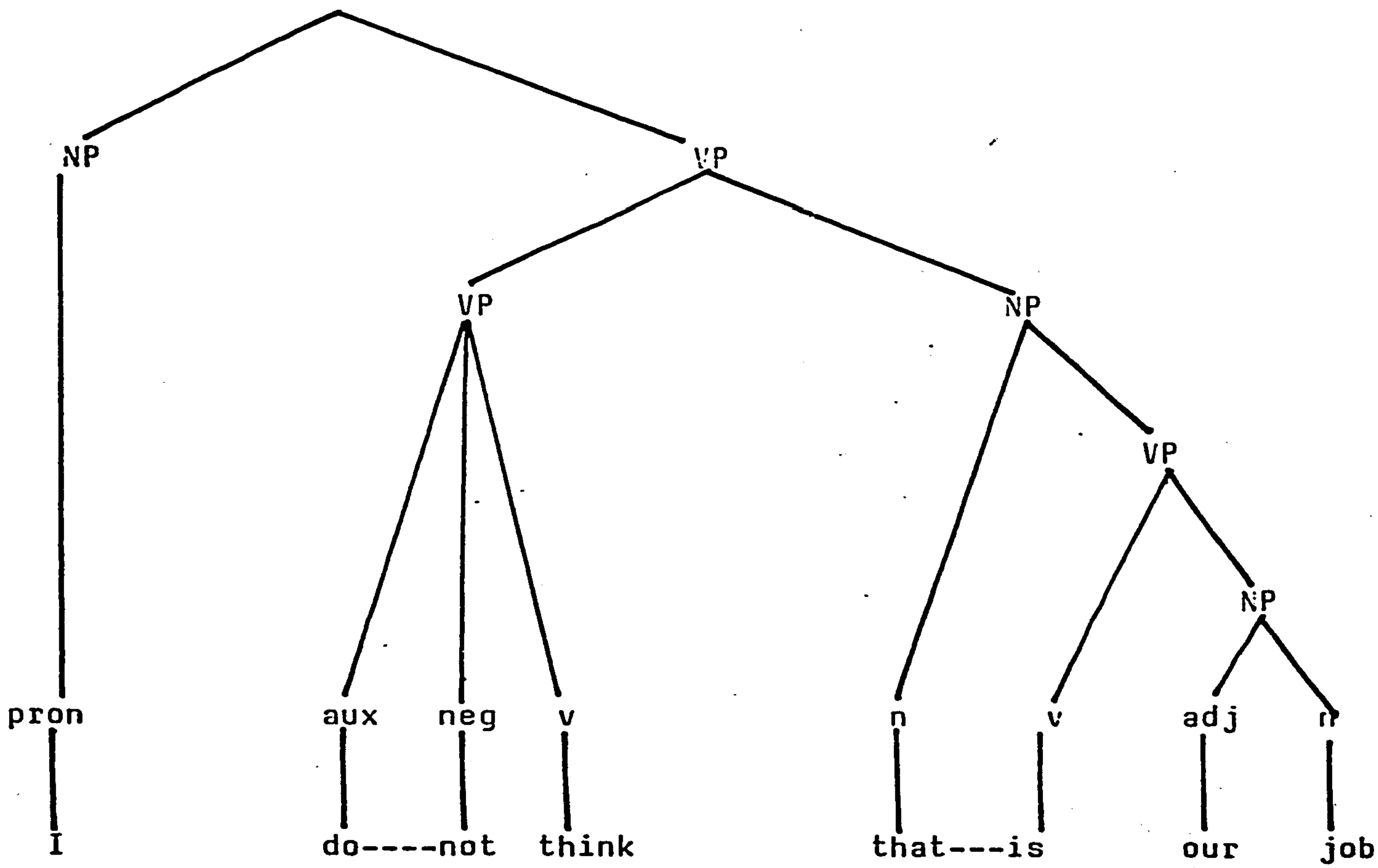

NP
VP
NP
NP
conj
NP
VP
VP
pron
v
art
n
pron
aux
v
n
That
is
the
mark
that
you
have
made
it

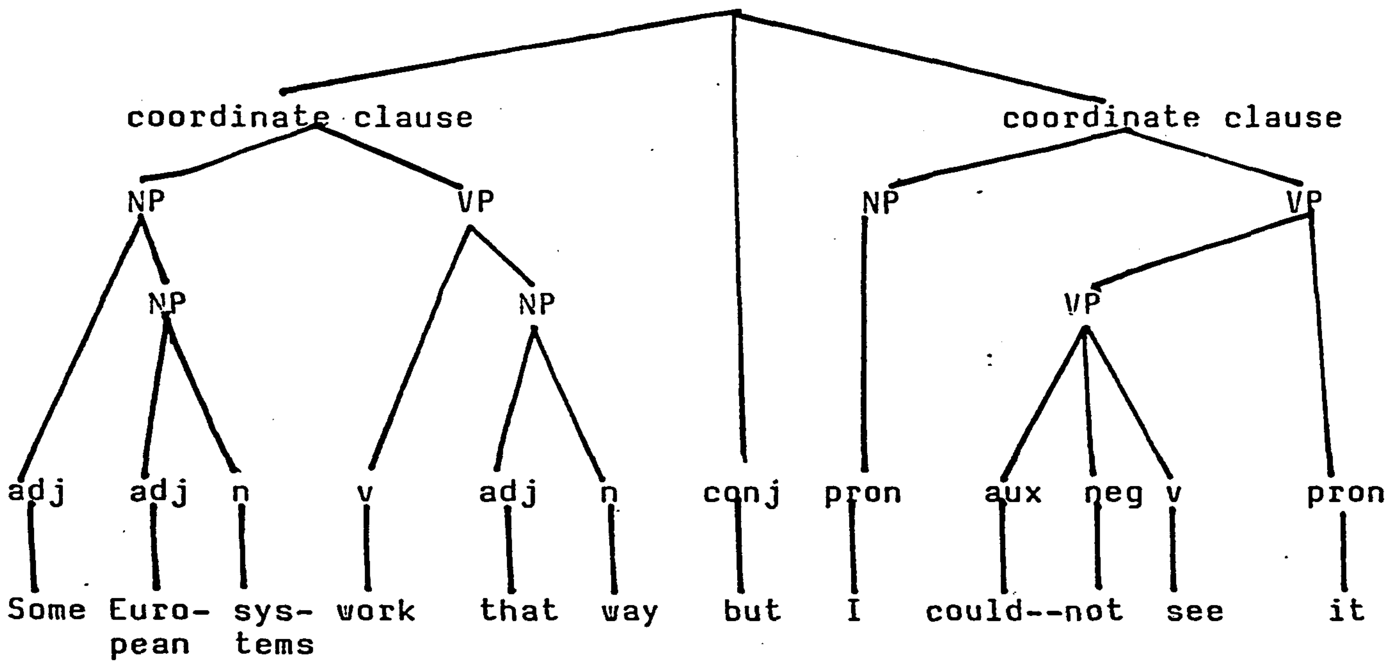
KM
14

KM
15

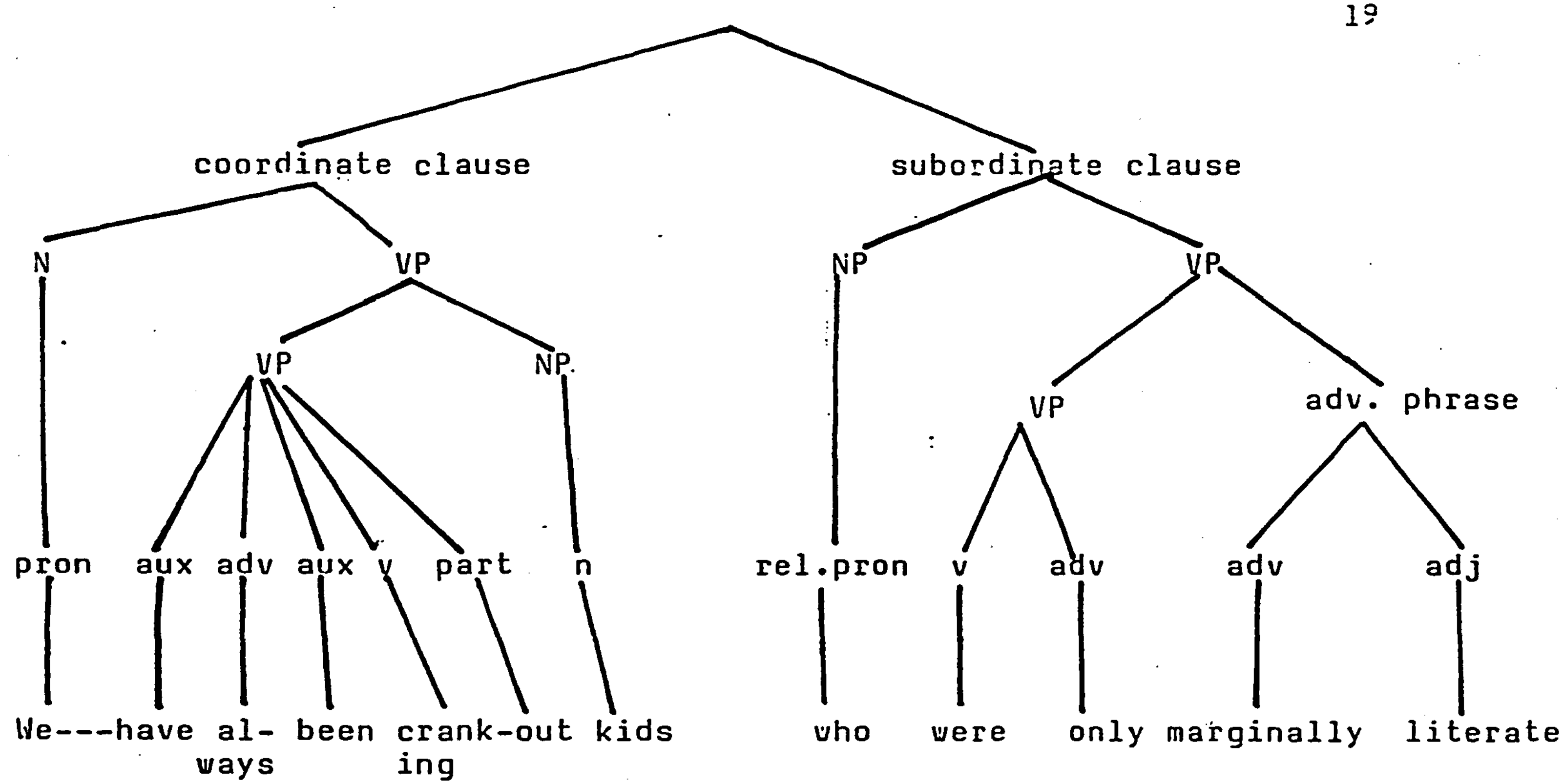

NP
VP
VP
NP
NP
PP
NP
pron
aux
neg
v
adj
n
prep
art
adj
n
n
I
can--not
get
that
excited
about
the
SAT
scores
falling

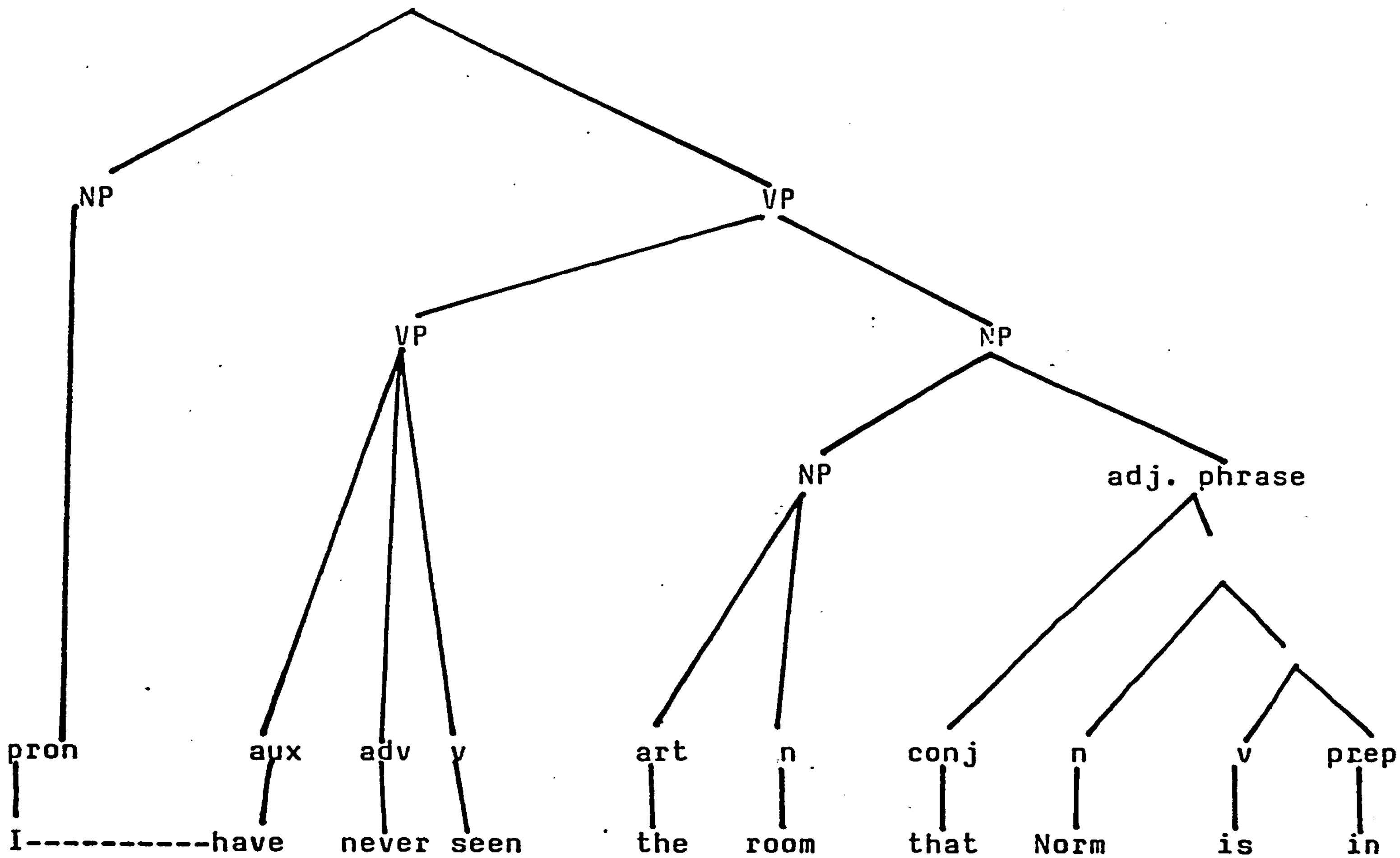
KM
16


NP
VP
VP
NP
VP
NP
pron
aux
neg
v
n
v
adj
n
I
do----not
think
that---is
our
job

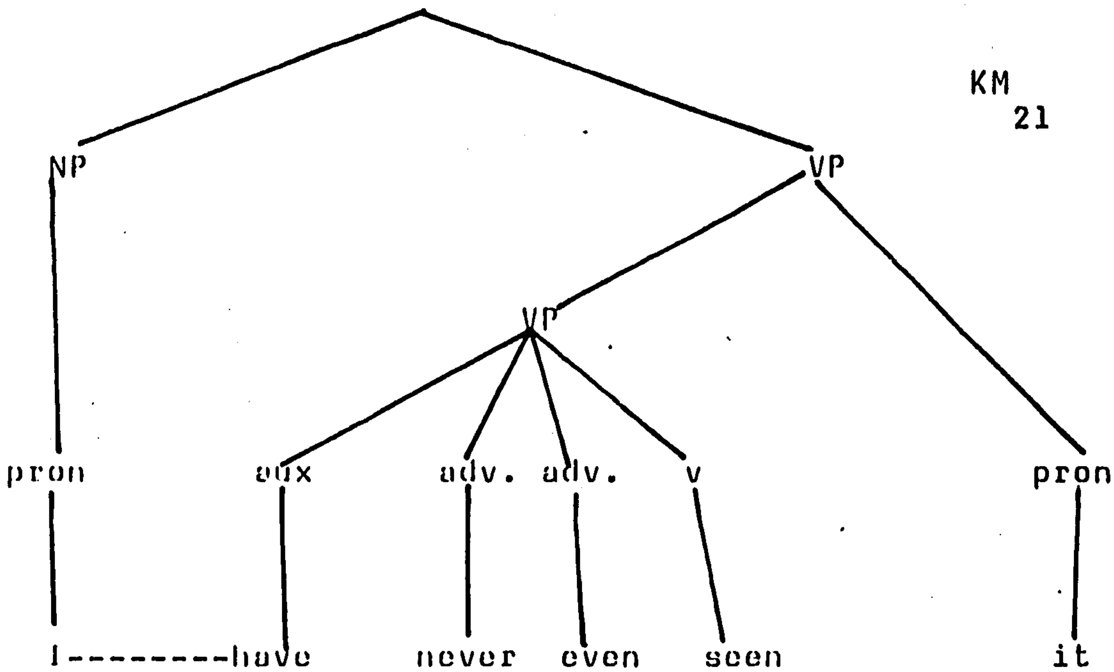

coordinate clause
coordinate clause
NP
VP
NP
VP
NP
NP
VP
adj
adj
n
v
adj
n
conj
pron
aux
neg
v
pron
Some
Euro-
pean
sys-
tems
work
that
way
but
I
could--not
see
it

KM
19

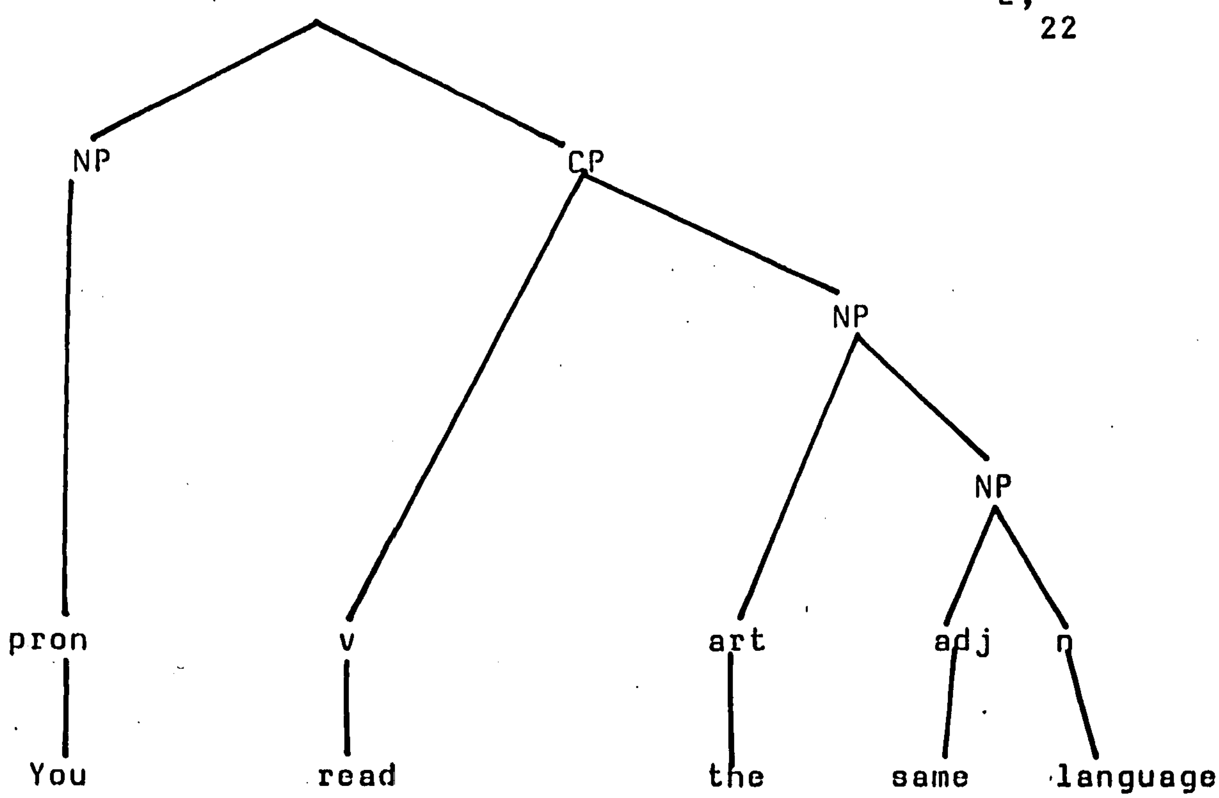

coordinate clause
subordinate clause
N
VP
NP
VP
VP
NP
VP
adv. phrase
pron
aux
adv
aux
v
part
n
rel.pron
v
adv
adv
adj
We---have
al-ways
been
crank-ing
out
kids
who
were
only
marginally
literate

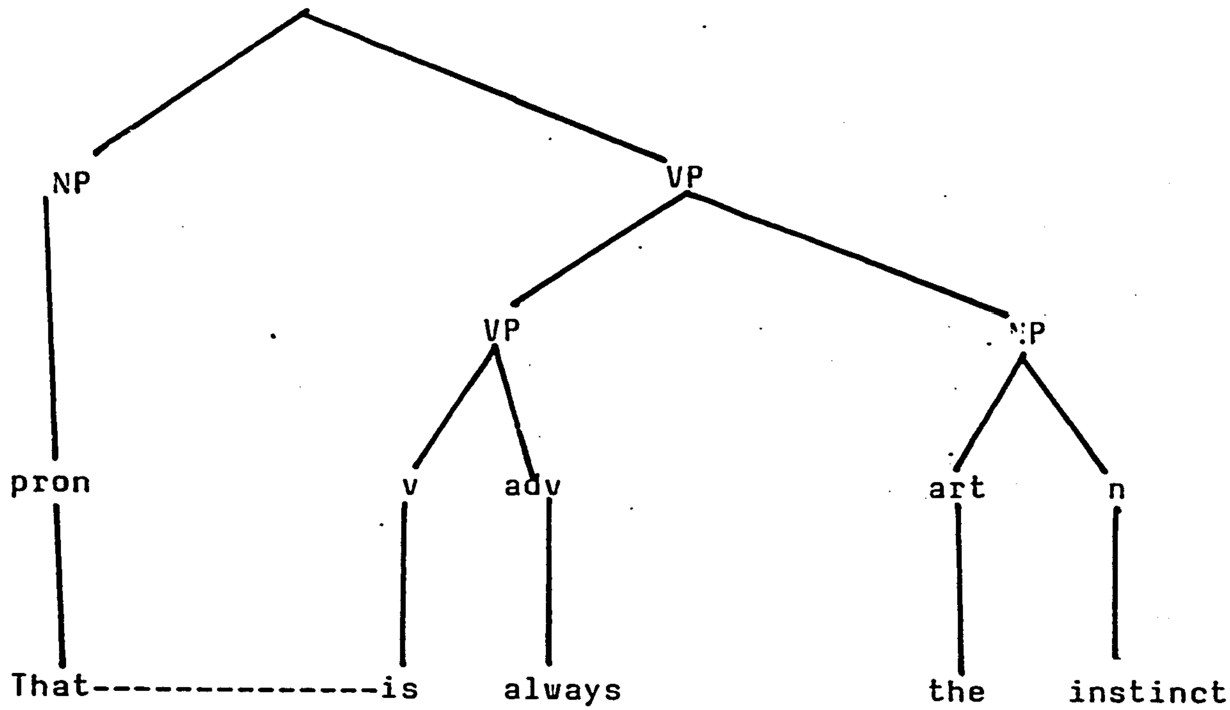

NP
VP
VP
NP
NP
adj. phrase
pron
aux
adv
v
art
n
conj
n
v
prep
I----------have never seen
the room
that Norm is in

KM
21

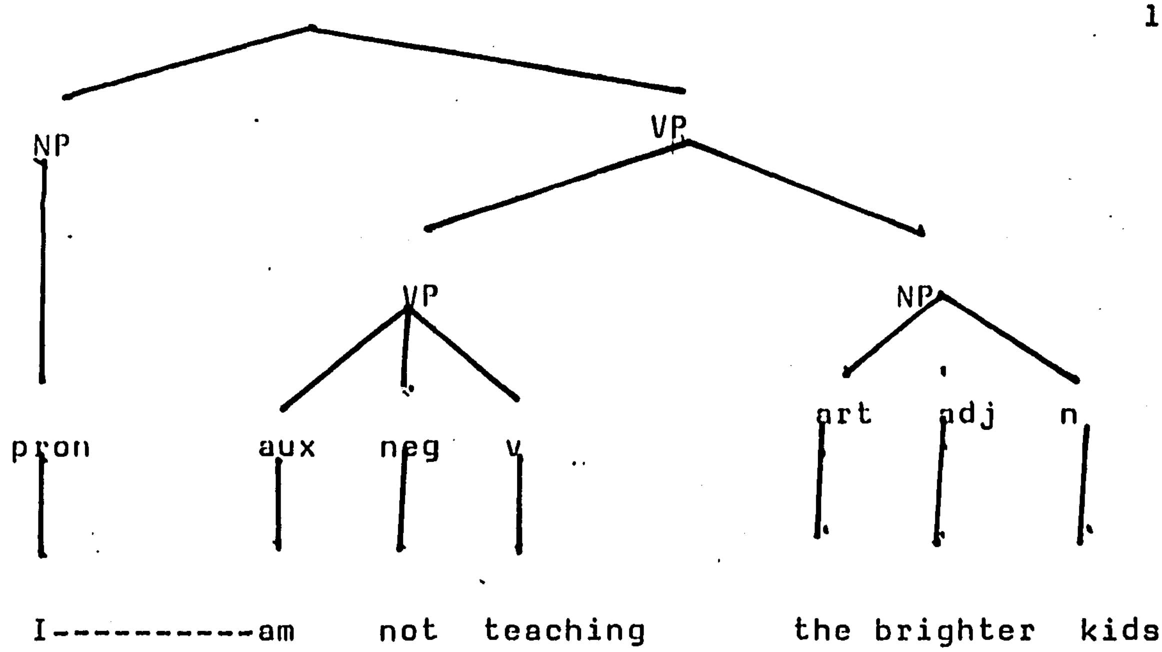

NP
VP
VP
pron
aux
adv.
adv.
v
pron
I--------have
never
even
seen
it

L,
22

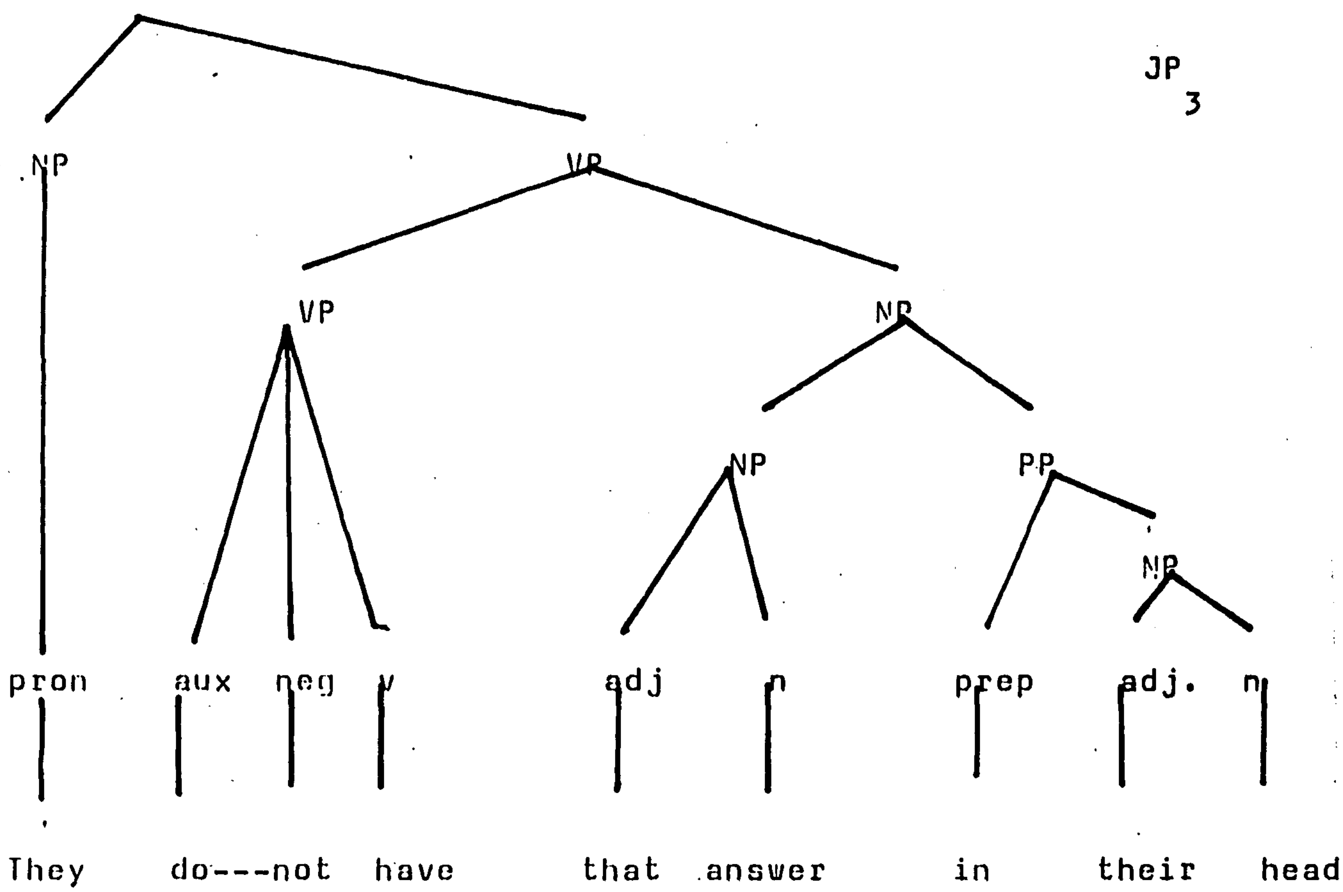

NP
CP
NP
NP
pron
v
art
adj
n
You
read
the
same
language


NP
VP
VP
NP
pron
v
adv
art
n
That--------------is
always
the
instinct

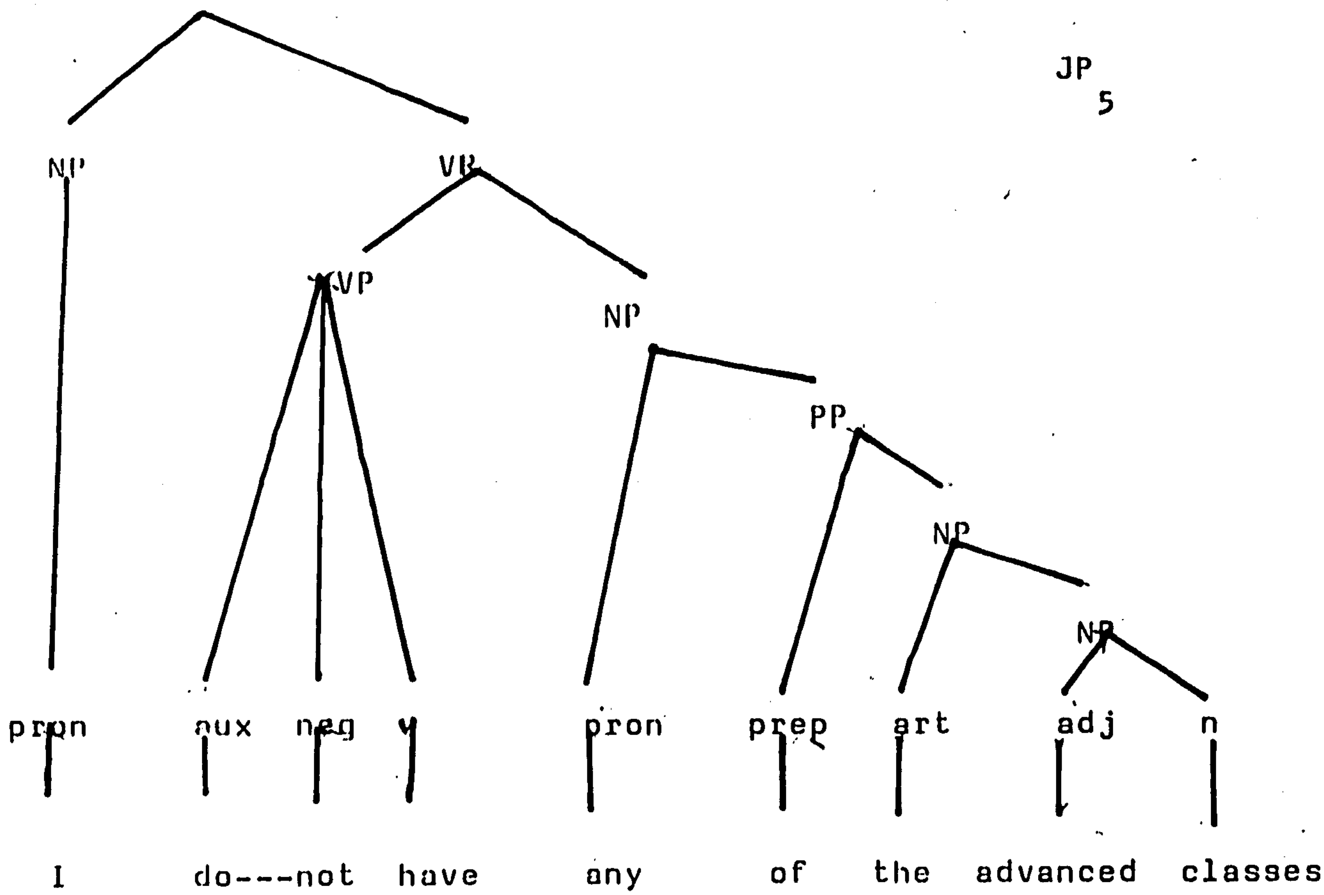
JP $_1$


NP
VP
VP
NP
pron
aux
neg
v
art
adj
n
I----------am
not
teaching
the brighter
kids

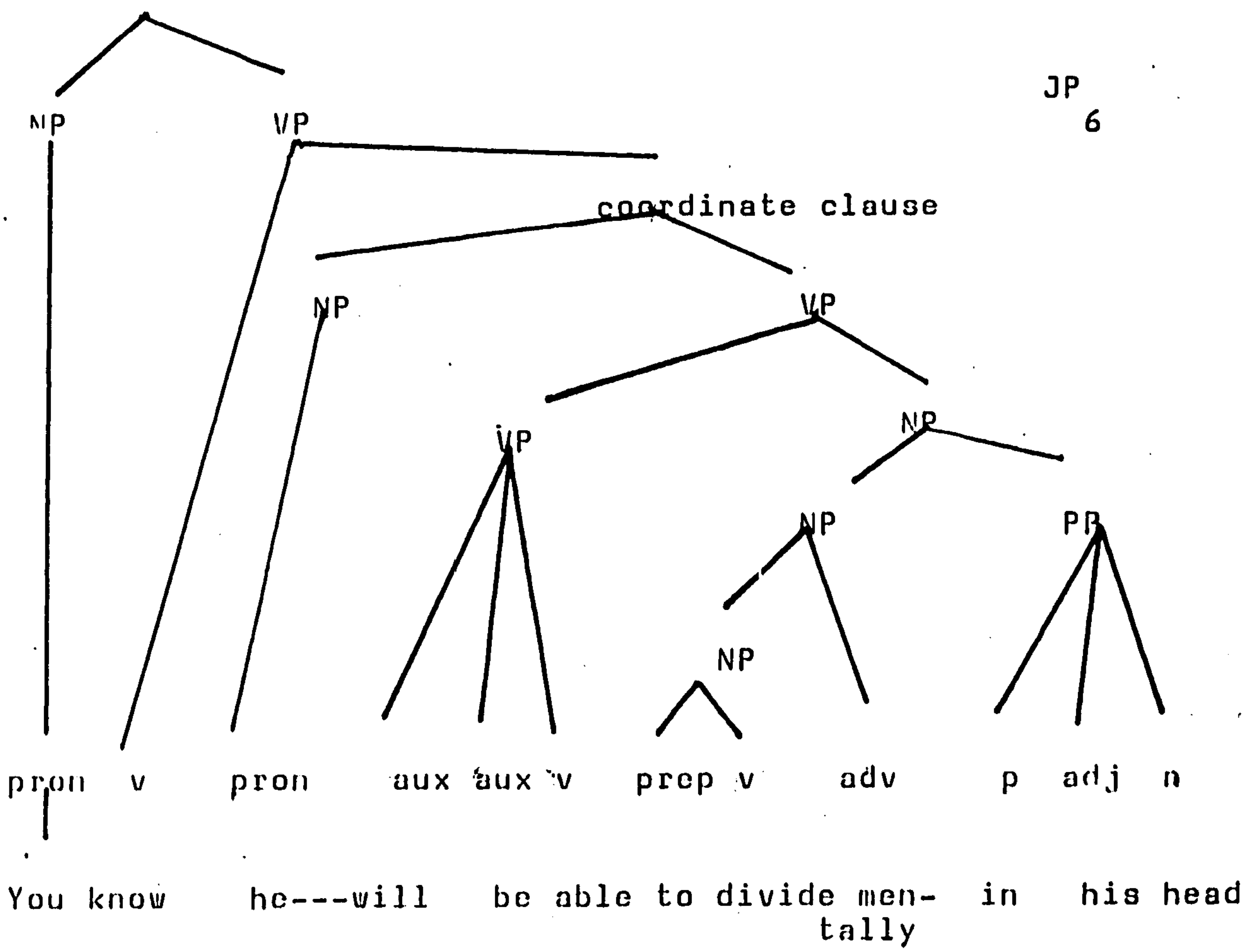
JP $_3$


NP
VP
VP
NP
NP
PP
NP
pron
aux
neg
v
adj
n
prep
adj.
n
They
do---not
have
that
answer
in
their
head

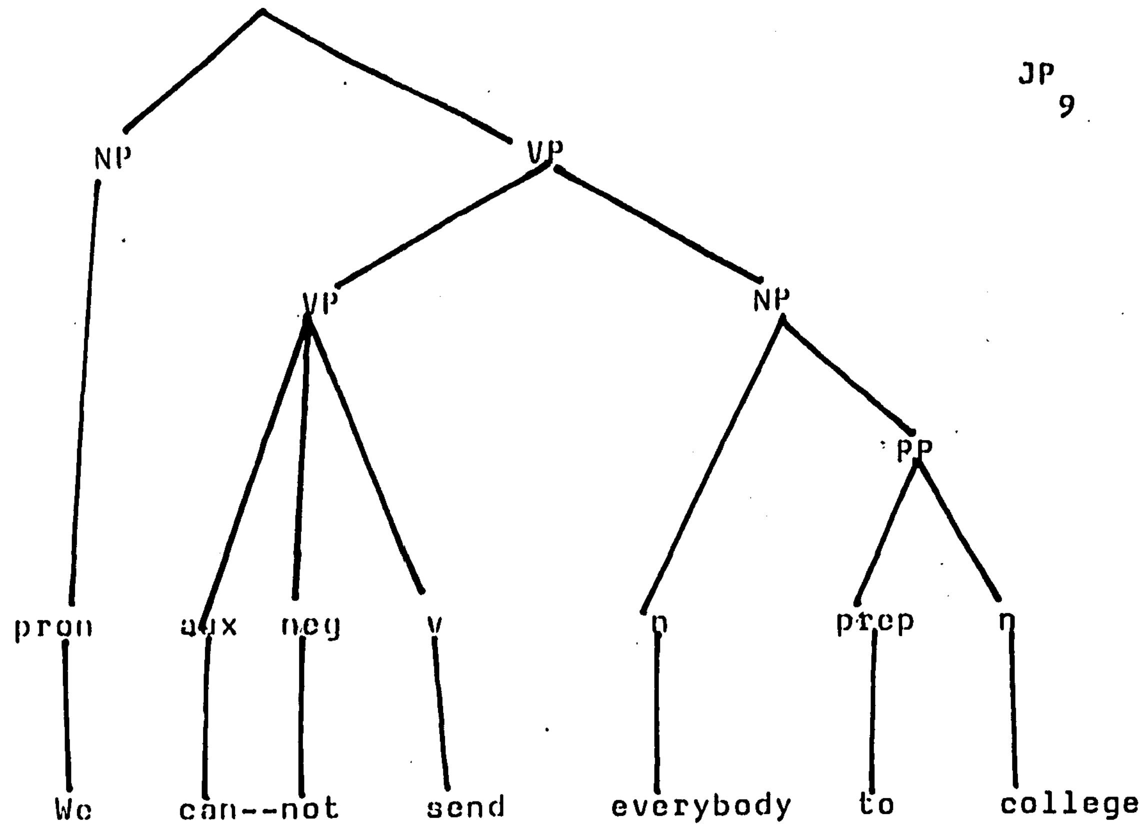
JP5


NP
VP
VP
NP
PP
NP
NP
pron
aux
neg
v
pron
prep
art
adj
n
I
do---not
have
any
of
the
advanced
classes

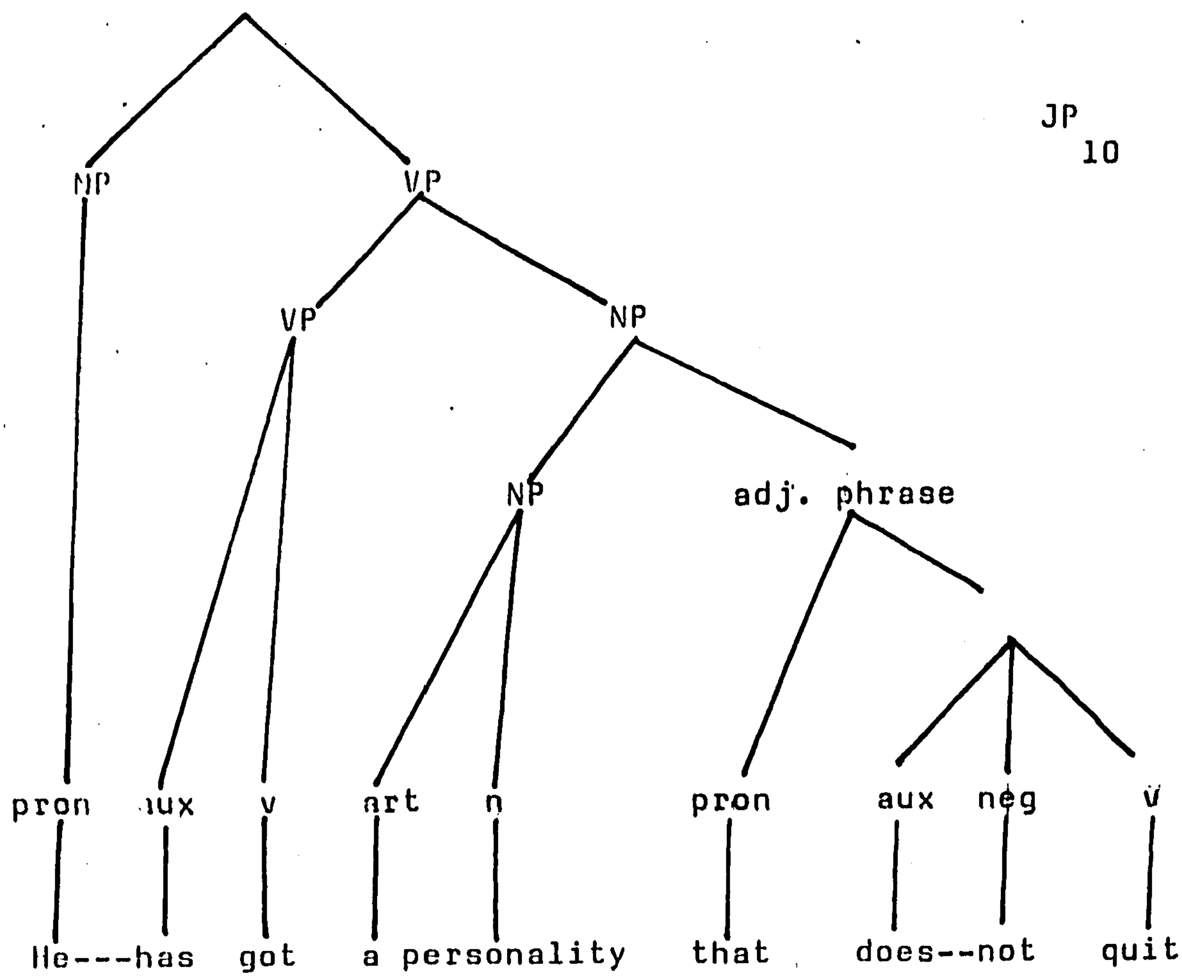
JP6


NP
VP
coordinate clause
NP
VP
VP
NP
NP
PP
NP
pron
v
pron
aux
aux
v
prep
v
adv
p
adj
n
You know
he---will
be able to divide men-
tally
in
his head

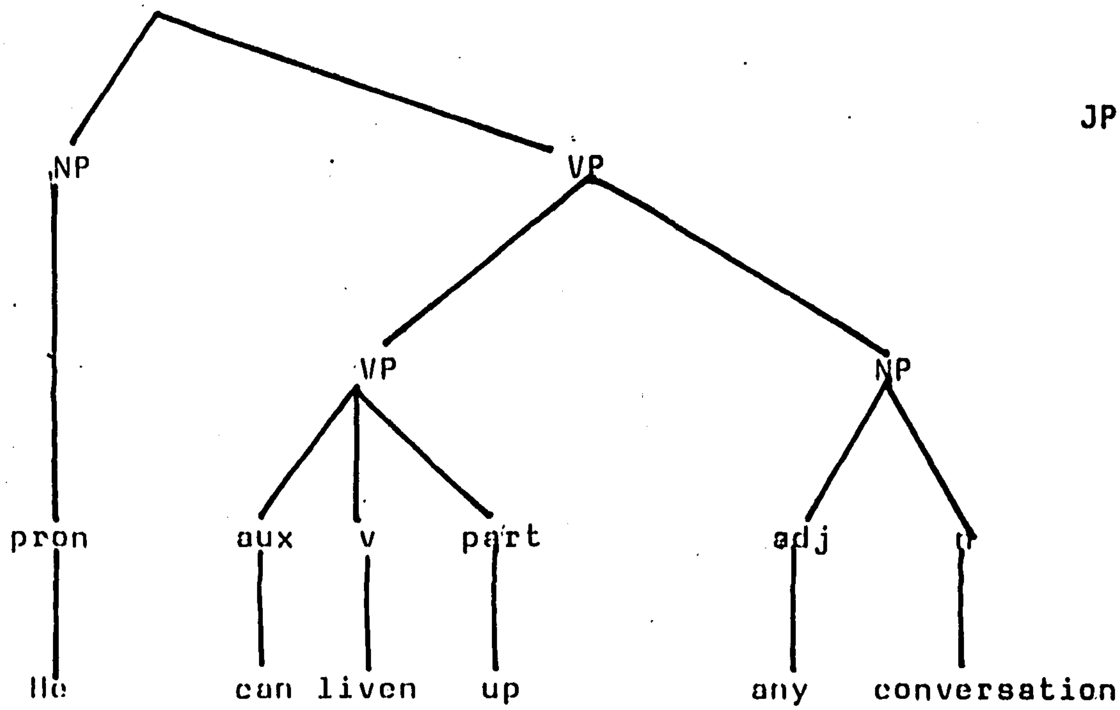
JP_9


NP
VP
VP
NP
PP
pron
aux
neg
v
n
prep
n
We
can--not
send
everybody
to
college

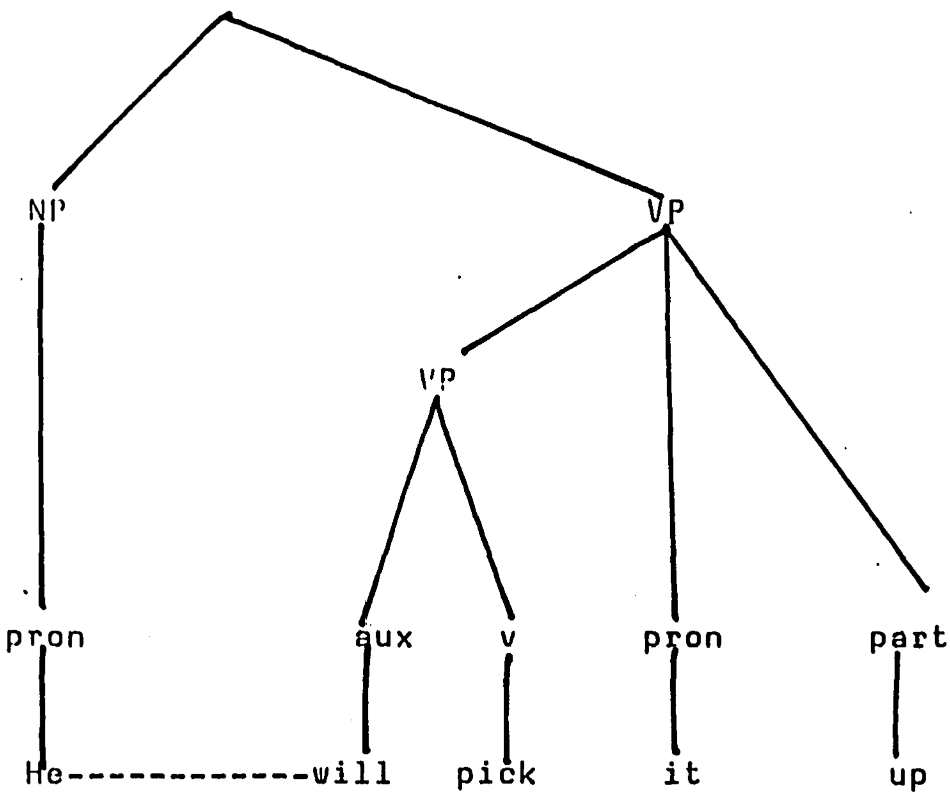
JP_{10}


NP
VP
VP
NP
NP
adj. phrase
pron
aux
v
art
n
pron
aux
neg
v
He---has
got
a personality
that
does--not
quit

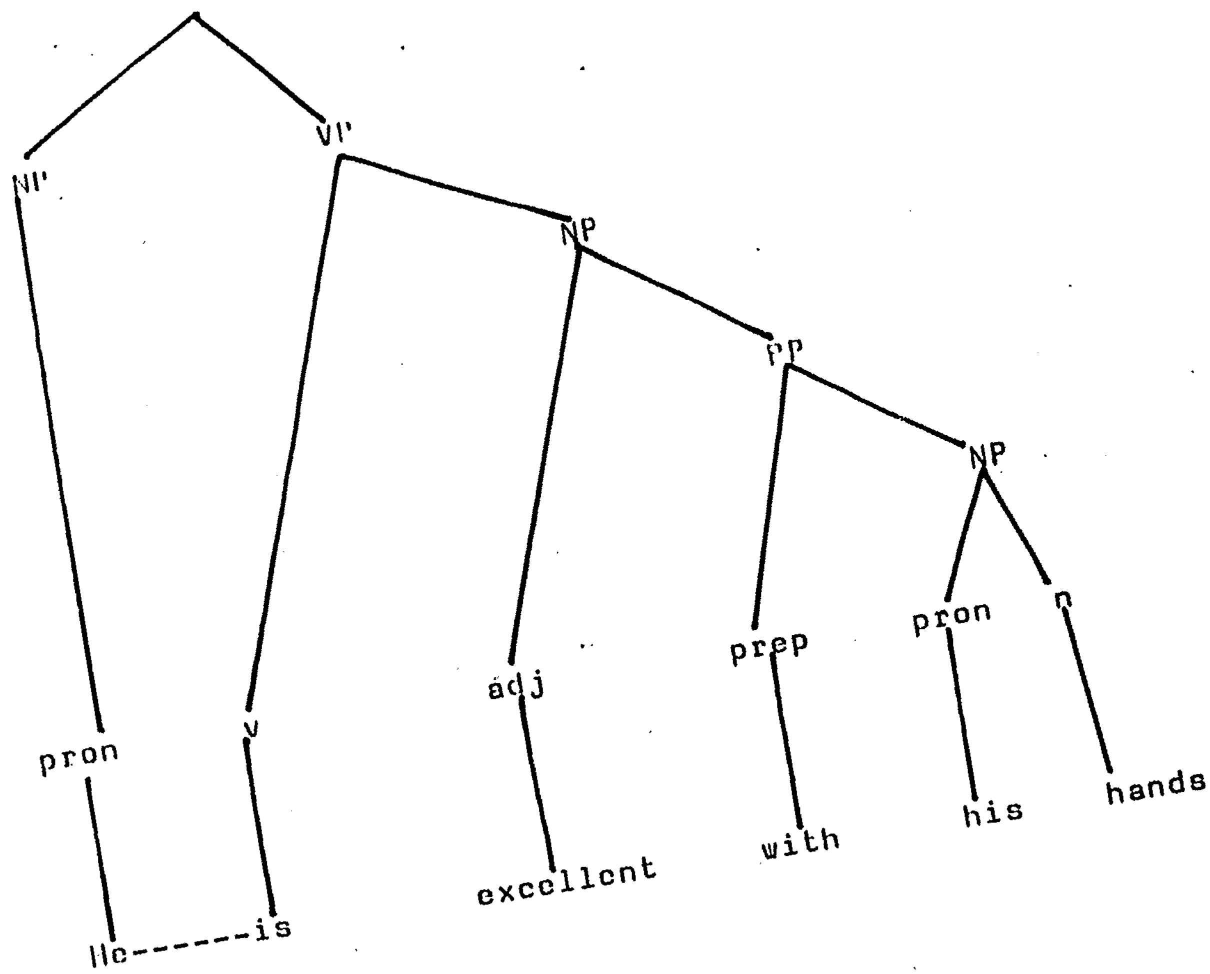
JP 11


NP
VP
VP
NP
pron
aux
v
part
adj
He
can
liven
up
any
conversation

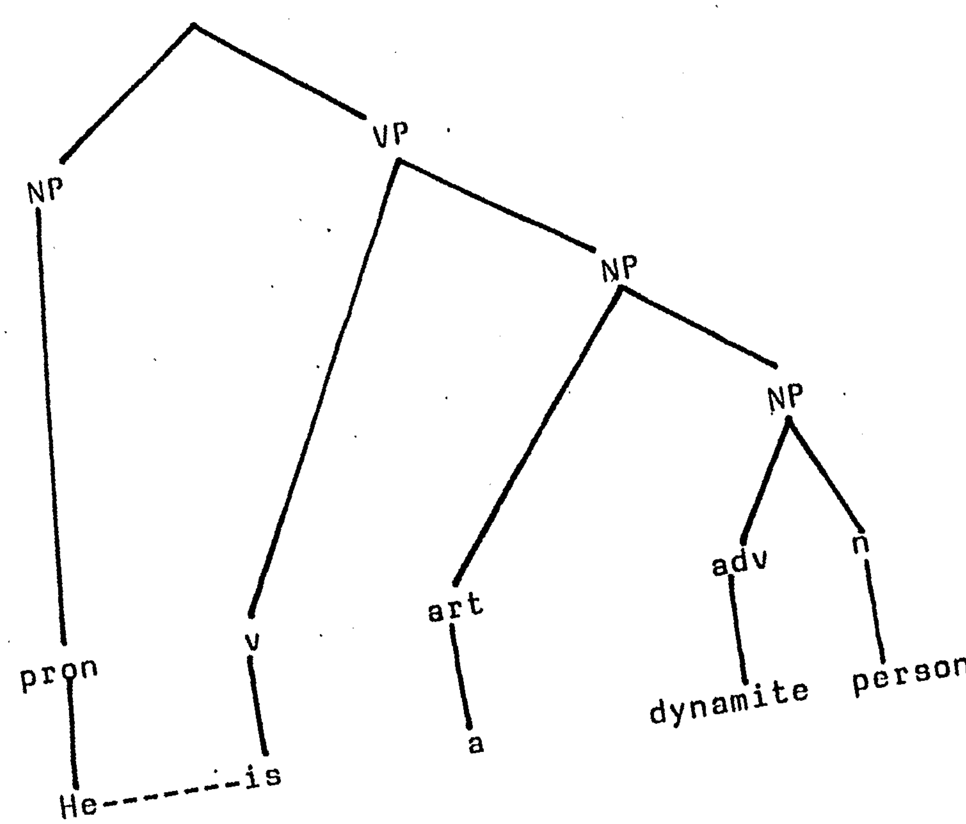
JP 16


NP
VP
VP
pron
aux
v
pron
part
He------------will
pick
it
up

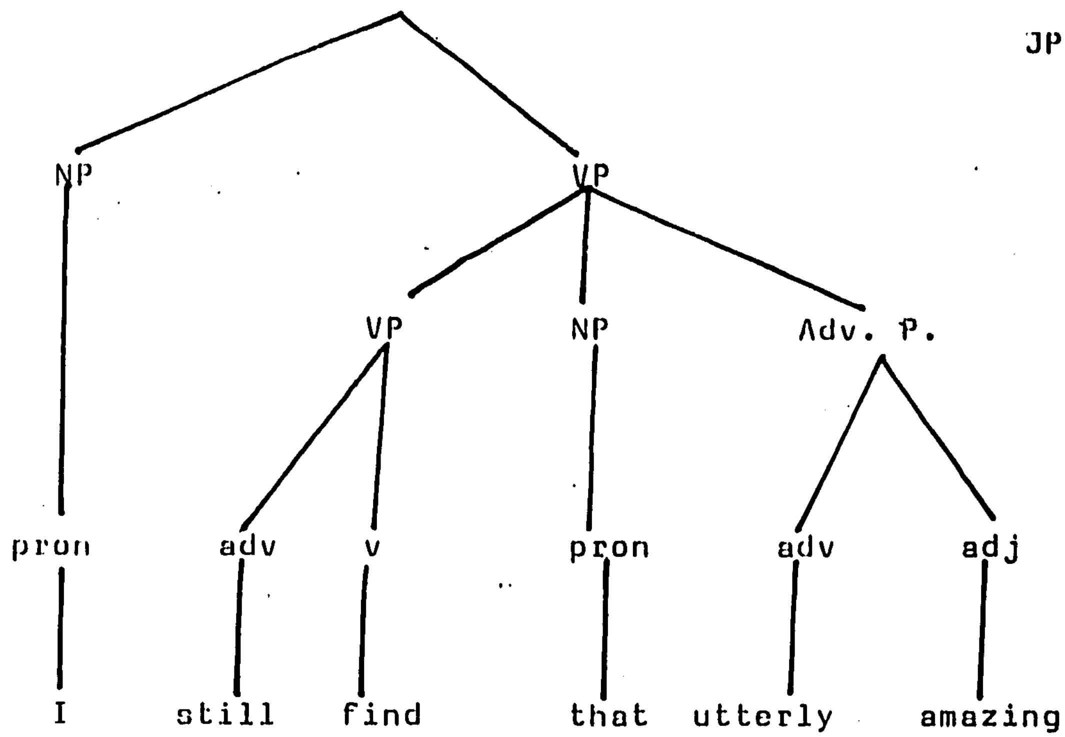
JP 19


NP
VP
NP
PP
NP
pron
v
adj
prep
pron
n
He------is
excellent
with
his
hands

JP 20


NP
VP
NP
NP
pron
v
art
adv
n
He-------is
a
dynamite
person

JP 22


NP
VP
VP
NP
Adv. P.
pron
adv
v
pron
adv
adj
I
still
find
that
utterly
amazing

APPENDIX B

INSTRUCTIONS TO LISTENERS FOR S/R IDENTIFICATION TASK

In this experiment you will be listening to sentences spoken by three different speakers. Some of these sentences were spoken spontaneously in conversation while others were read from a script. Listen carefully to each sentence. If you think the sentence was read, circle R. If you think it was spoken spontaneously, circle S. On the response sheet there are three different size Rs and three different size Ss after each sentence. In indicating your response, circle the largest letter if you are positive about your choice. If you are fairly certain, but not completely confident in your response, circle the middle size letter. If you have considerable doubt about your response circle the smallest letter. You must make a choice. Guess if necessary.

You will hear a hissing sound at the beginning of each sentence trial. After the hiss the sentence will be presented twice. Following the second presentation there will be a brief silent interval during which you can mark your response sheet.

These sentences were computer manipulated and may sound strange. Do not worry about the content of the sentences; just concentrate on deciding whether they were read or spoken spontaneously.

Would you also please answer the question on the bottom of the last response sheet?

Thank you for your time.

APPENDIX C

SAMPLE RESPONSE SHEET

Listener's Name________________________________ Date______________

Speaker #1

1. Some European systems work that way, but I couldn't see it. R R R S S S
2. No, he couldn't. R R R S S S
3. Some European systems work that way, but I couldn't see it. R R R S S S
4. Some European systems work that way, but I couldn't see it. R R R S S S
5. Some European systems work that way, but I couldn't see it. R R R S S S
6. No, he couldn't. R R R S S S
7. No, he couldn't. R R R S S S
8. No, he couldn't. R R R S S S
9. And the others found jobs where they could be trained. R R R S S S
10. You read the same language. R R R S S S
11. I've never seen the room that Norm's in. R R R S S S
12. And you wonder how they're going to make it. R R R S S S

13. I can't get that excited about the SATs scores falling. R R R S S S

14. That's the mark that you've made it. R R R S S S

15. There is no New York Written accent. R R R S S S

16. Not one of them had to do with teaching. R R R S S S

17. This is an aptitude test. R R R S S S

18. I've never even seen it. R R R S S S

19. They tend to lose things literally. R R R S S S

20. They're not as healthy. R R R S S S

21. I don't think that's our job. R R R S S S

22. It measures innate ability. R R R S S S

23. But they're not the ones taking the SATs. R R R S S S

24. It's almost a psychological thing. R R R S S S

25. That's always the instinct. R R R S S S

26. We've always been cranking out kids who were only marginally literate. R R R S S S

27. They're losers. R R R S S S

Speaker #2

1. I should be, but I'm not. RRRSSs

2. I should be, but I'm not. RRRSSs

3. If you're satisfied with it, then you are developing yourself. RRRSSs

4. If you're satisfied with it, then you are developing yourself. RRRSSs

5. I should be, but I'm not. RRRSSs

6. I should be, but I'm not. RRRSSs

7. If you're satisfied with it, then you are developing yourself. RRRSSs

8. If you're satisfied with it, then you are developing yourself. RRRSSs

9. You said a Tuesday would be good. RRRSSs

10. Next Tuesday's no good. RRRSSs

11. Something that they need. RRRSSs

12. That's effective communication. RRRSSs

13. But I will start to do that shortly. RRRSSs

14. Feminism. RRRSSs

15. This is just what men are afraid of. RRRSSs

16. I know of a good Italian restaurant in Plainfield.

RRRSSs

17. No.

18. That's between you and me and the microphone. RRRSSS

19. You don't know who contributed how much. RRRSSS

20. Everybody's freedom might be considered a threat to someone else. RRRSSS

21. Norrie. RRRSSS

22. I specifically asked her about articulation. RRRSSS

23. It has been suggested to me that I spend more time on that. RRRSSS

24. Maybe you should try it in your sneakers first. RRRSSS

25. I have a list of things I won't do. RRRSSS

26. I'll have to make some positive plans. RRRSSS

27. You'll help me do that. RRRSSS

28. Yeah. RRRSSS

29. This is what frightens men. RRRSSS

<u>Speaker #3</u>

1. In a graduate class one professor said something that made a lot of sense to me. RRRSSS

2. He says, "You know someone's got to be a garbageman." RRRSSS

3. He says, "You know someone's got to be a garbageman." R R R S S S

4. He says, "You know someone's got to be a garbageman." R R R S S S

5. In a graduate class one professor said something that made a lot of sense to me. R R R S S S

6. In a graduate class one professor said something that made a lot of sense to me. R R R S S S

7. In a graduate class one professor said something that made a lot of sense to me. R R R S S S

8. He says, "You know someone's got to be a garbageman." R R R S S S

9. But I have the average kids. R R R S S S

10. He's excellent with his hands. R R R S S S

11. Well, first of all let's look at it from the math angle. R R R S S S

12. You know he'll be able to divide mentally in his head. R R R S S S

13. They don't have that answer in their head. R R R S S S

14. He can liven up any conversation. R R R S S S

15. I'm not teaching the brighter kids. R R R S S S

16. In the sense that he doesn't sit down and read a book. R R R S S S

17. He's a dynamite person. R R R S S S

18. There is a lack of ability in just fundamental arithmetic. R R R S S s

19. I still find that utterly amazing. R R R S S s

20. I don't have any of the advanced classes. R R R S S s

21. But the slower kids can't do it, and they don't want to do it. R R R S S s

22. He's got a personality that doesn't quit. R R R S S s

23. Eddy Vent. R R R S S s

24. We can't send everybody to college. R R R S S s

25. He'll pick it up. R R R S S s

26. No. R R R S S s

What criteria did you use in making your decisions as to whether the sentences were read or were spoken spontaneously?

____

____

____

____

____

APPENDIX D

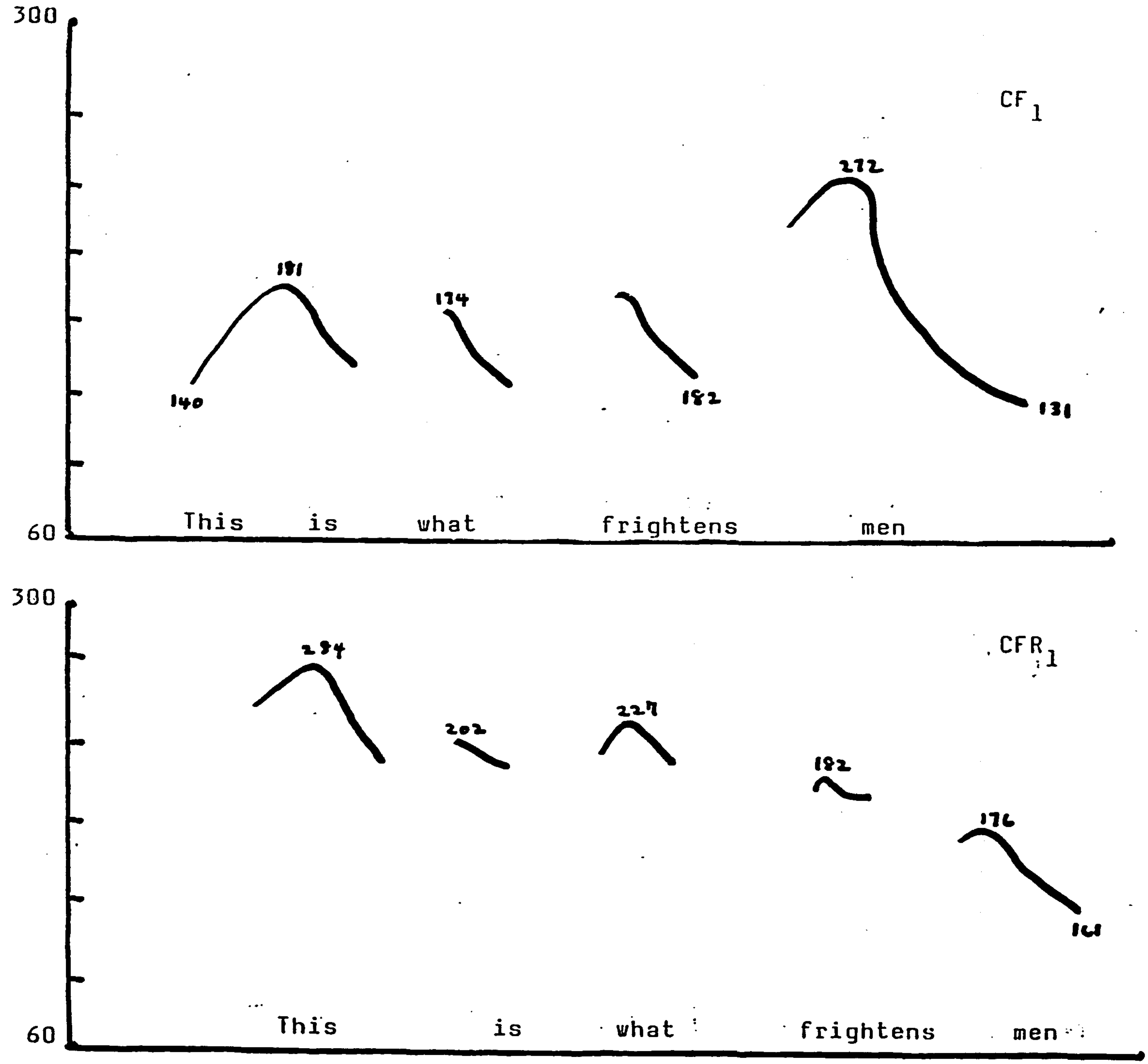
List of bisyllabic words used in making measurements for SUR

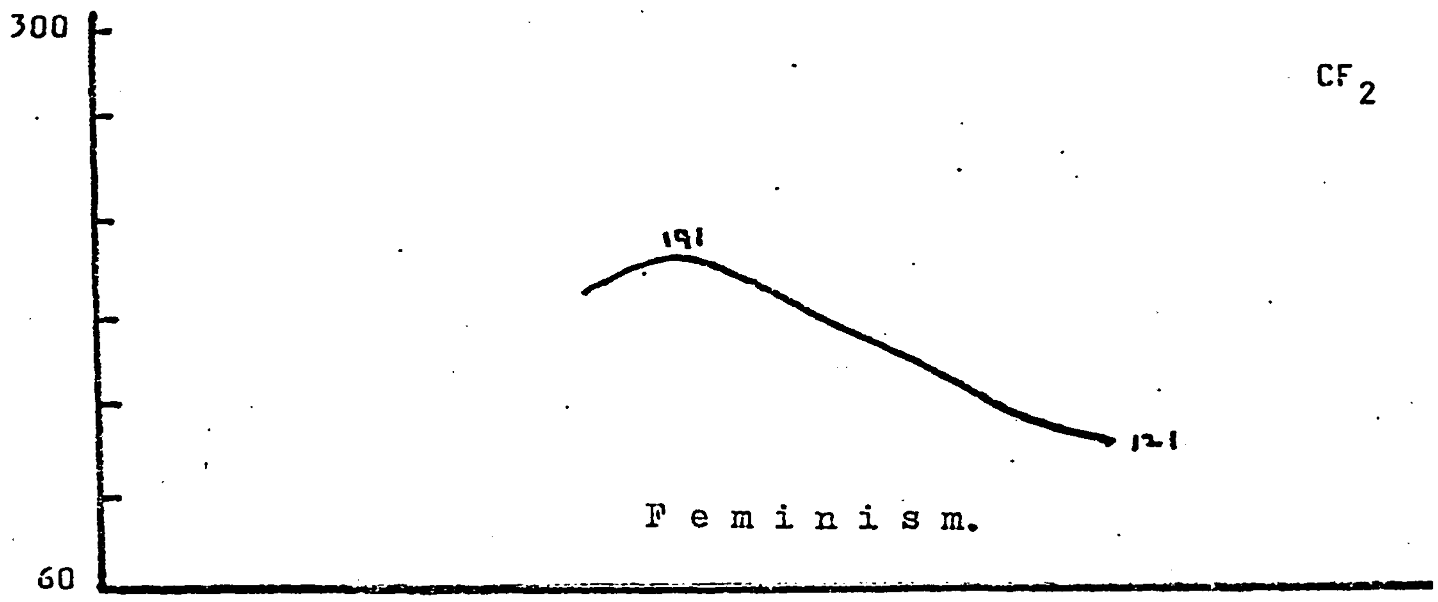
Speaker			
	frightens	afraid	freedon
	someone	yourself	assume
CF n=16	between	shortly	something
	Norrie	Tuesday	Tuesday's
	Chinese	Plainfield	maybe
	New York	written	accent
	losers	healthy	measures
	innate	almost	teaching
KM n=27	about	falling	systems
	couldn't (2)	always (2)	cranking
	only	never (2)	even
	language	instinct	others
	wonder	going	taking
	teaching	brighter	angle
	answer	advanced	classes
JP n=16	divide	slower	something
	college	doesn't (2)	liven
	someone's	garbage	person

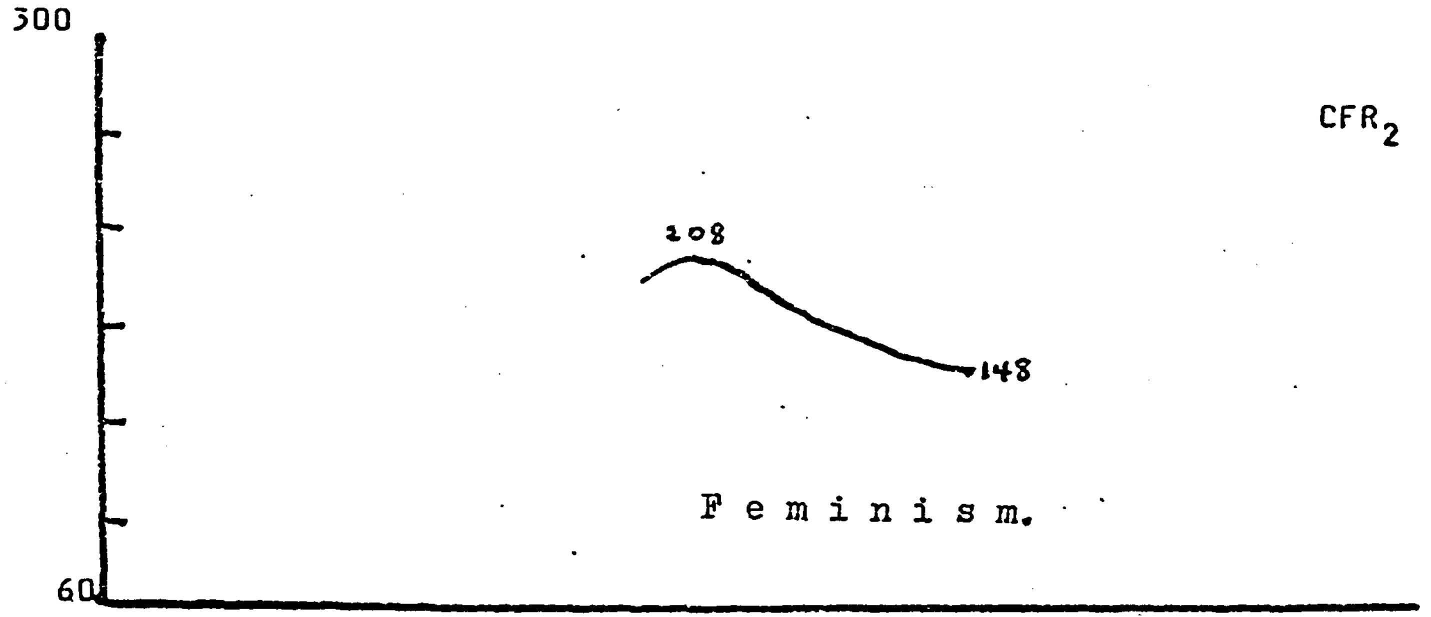
APPENDIX E

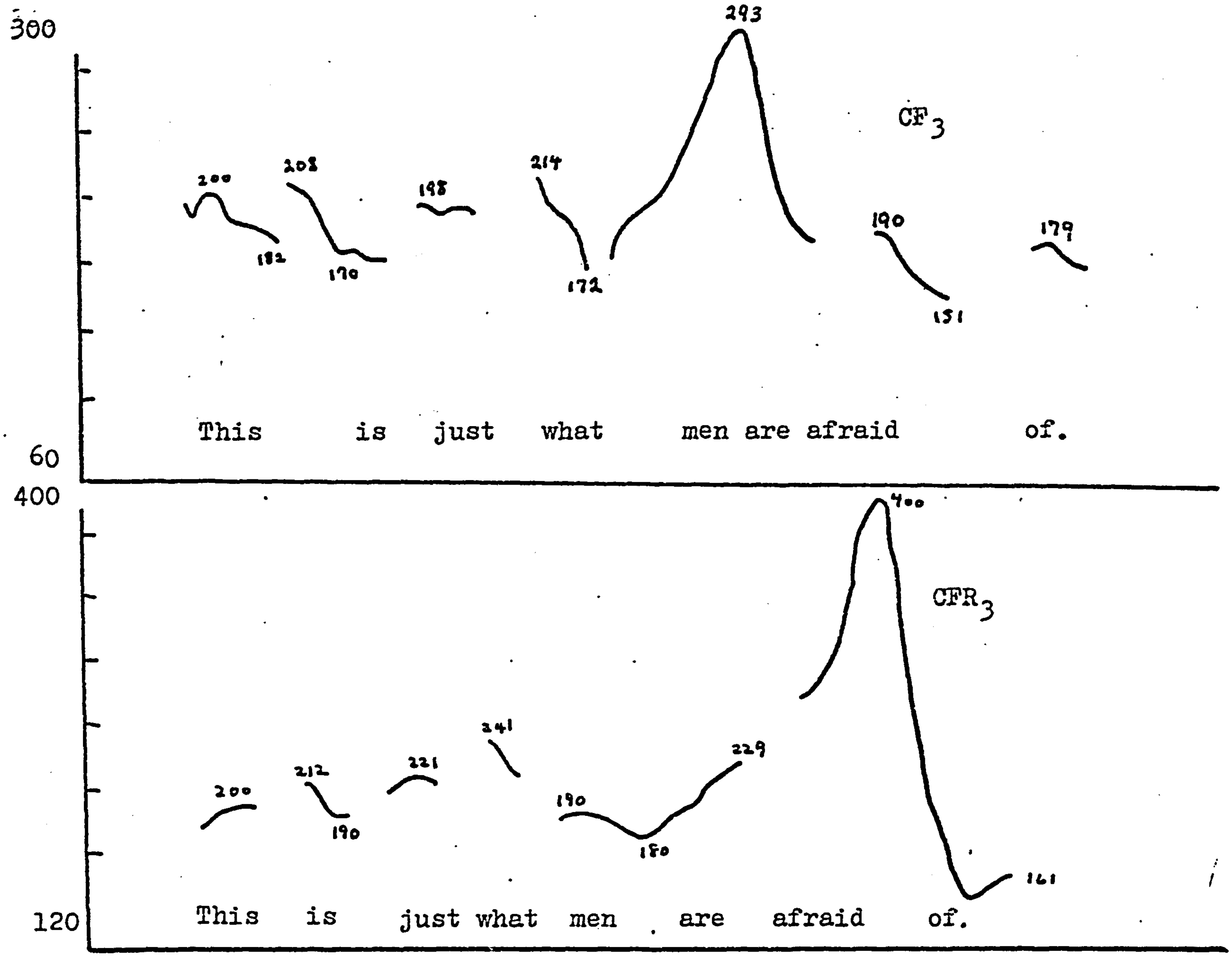
FUNDAMENTAL FREQUENCY CONTOURS

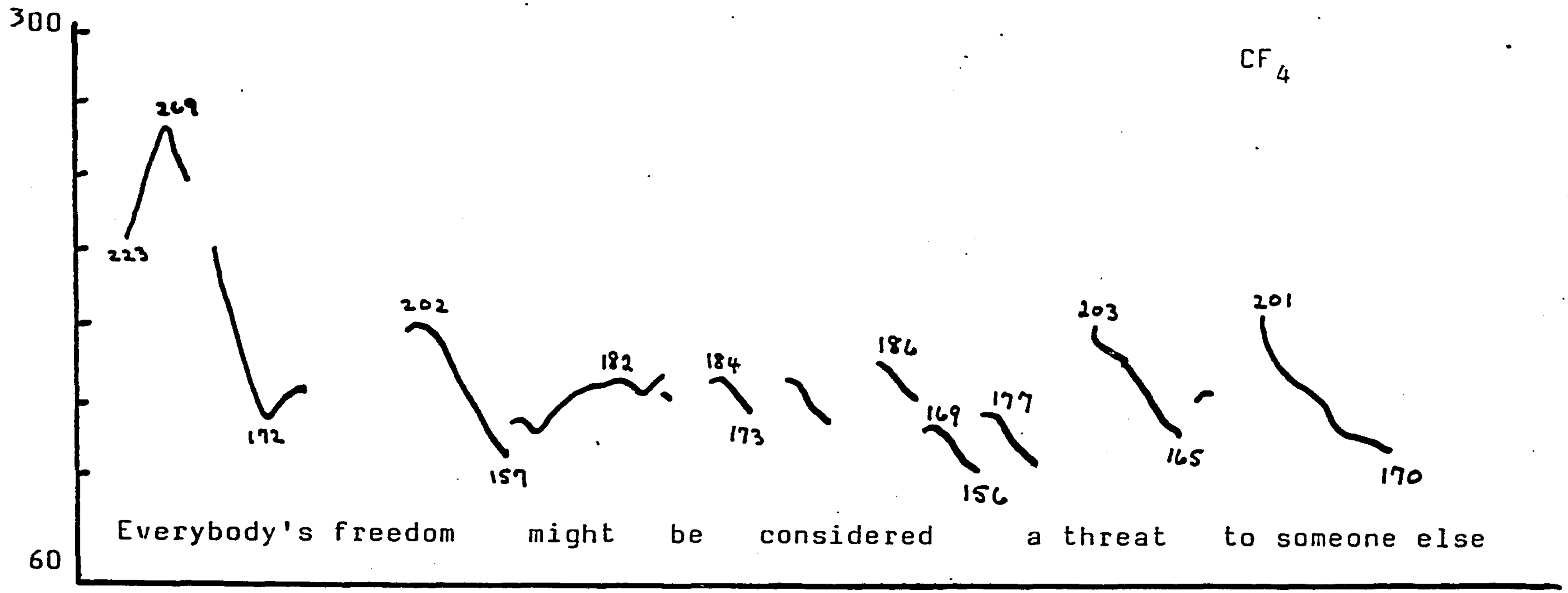
(Spontaneous Versions on top; read versions on bottom)

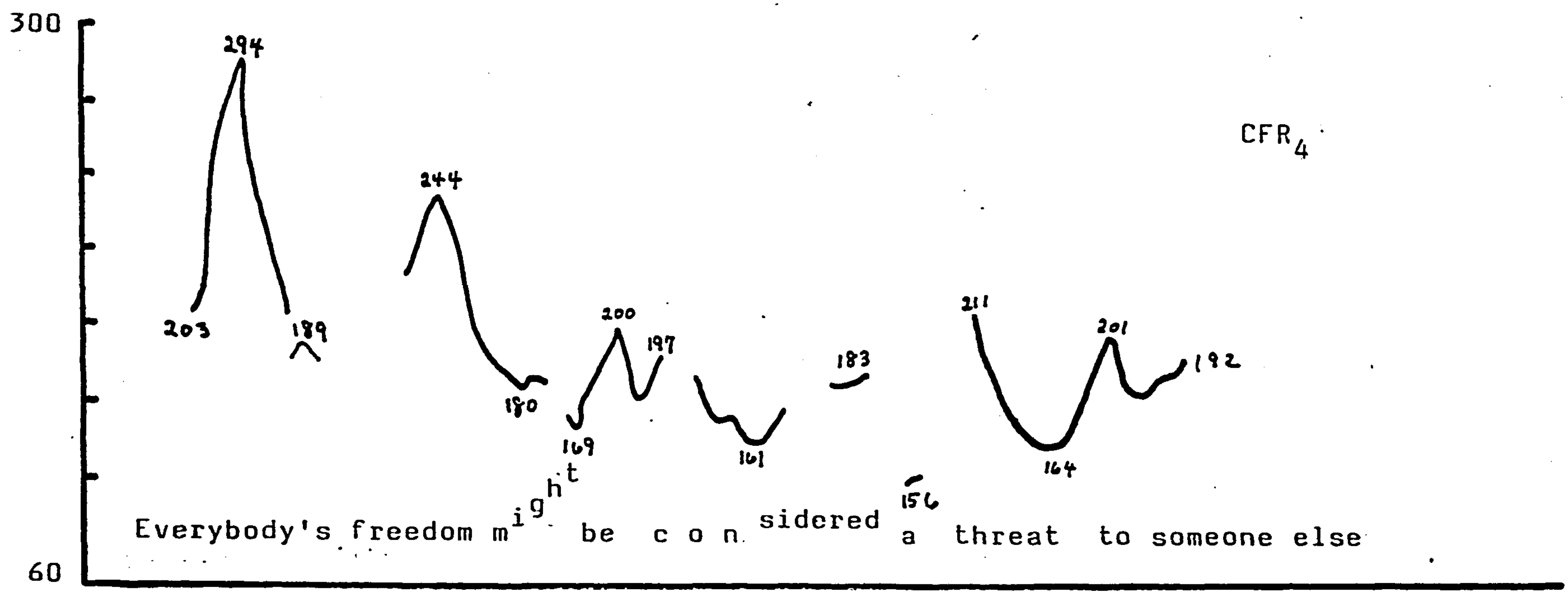


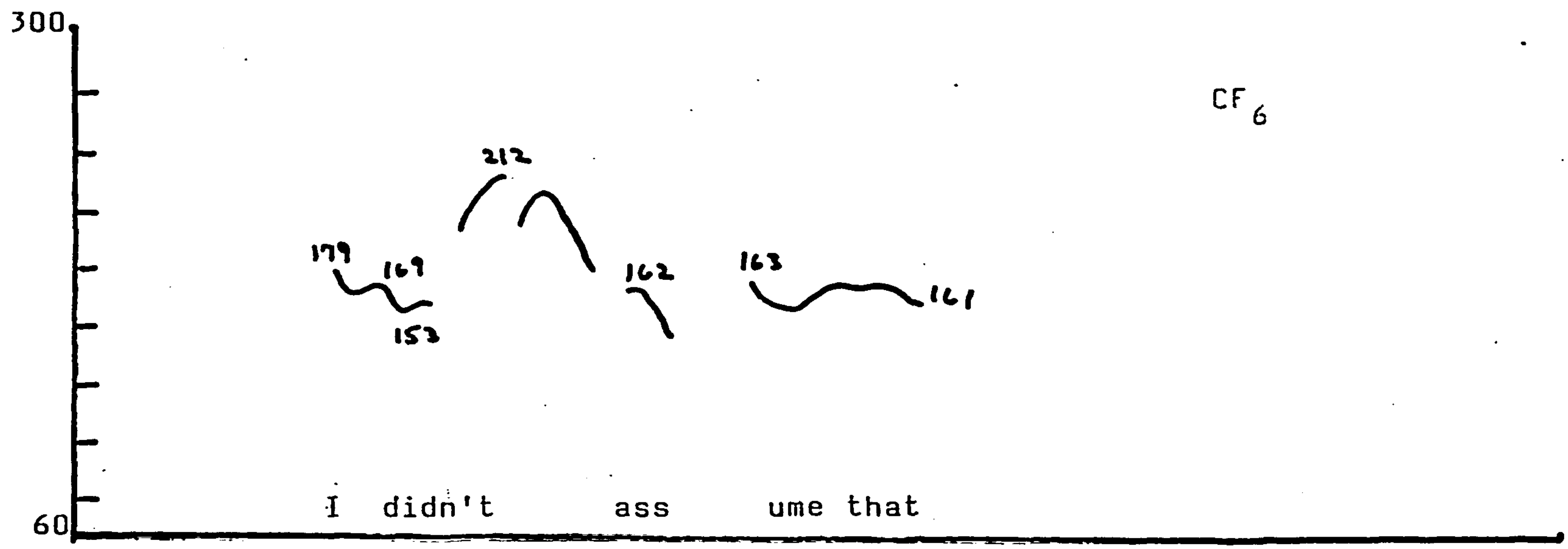

300
CF2
191
121
Feminism.
60

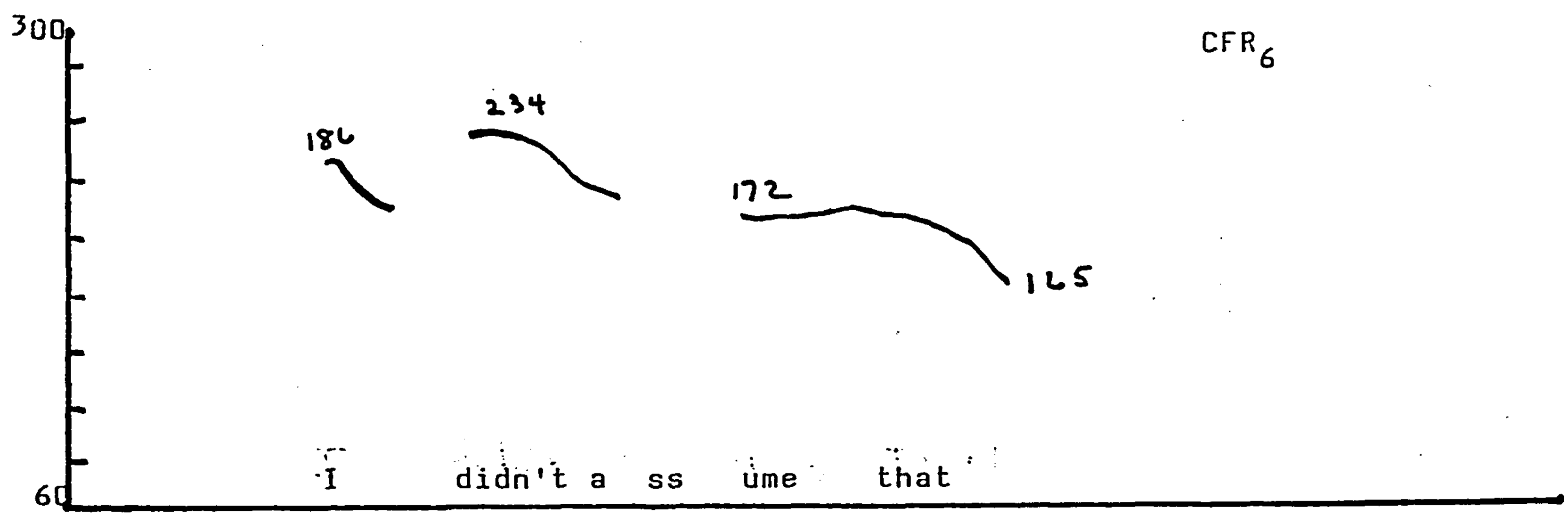

300
CFR2
208
148
Feminism.
60

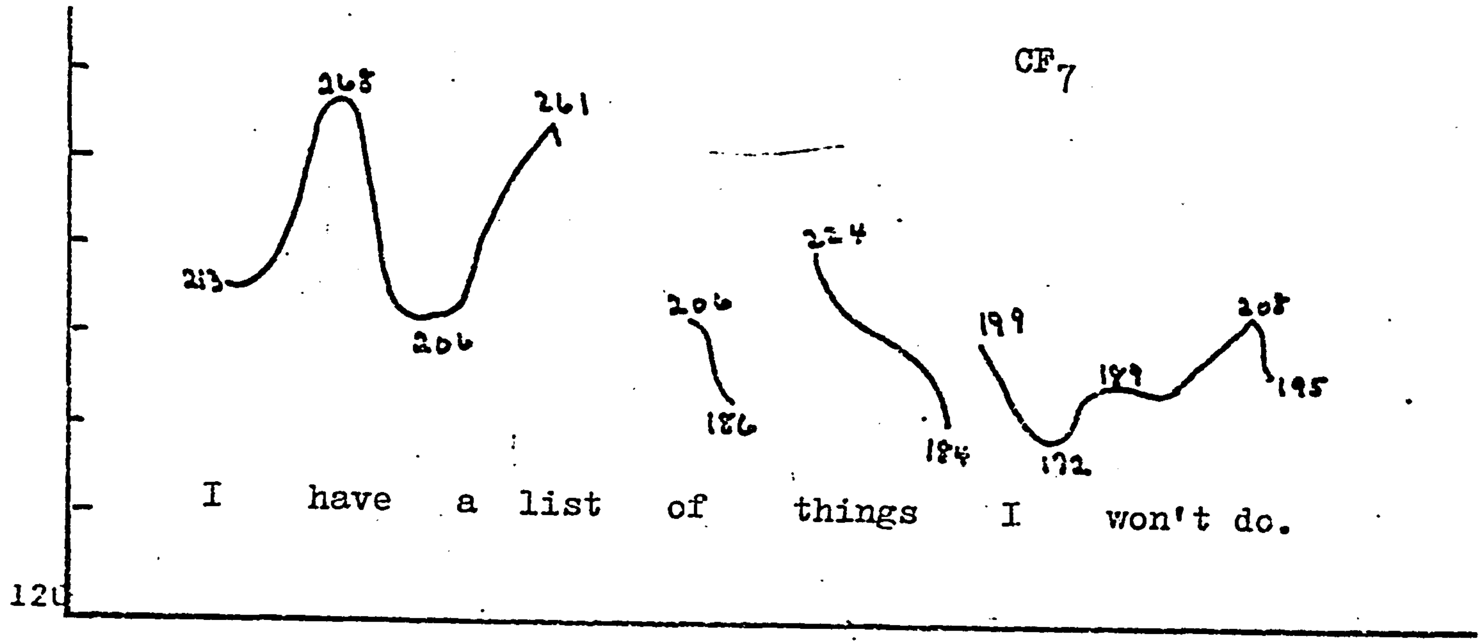

300
CF3
200
182
208
170
198
214
172
293
190
151
179
This is just what men are afraid of.
60
400
CFR3
200
212
190
221
241
190
180
229
400
161
This is just what men are afraid of.
120

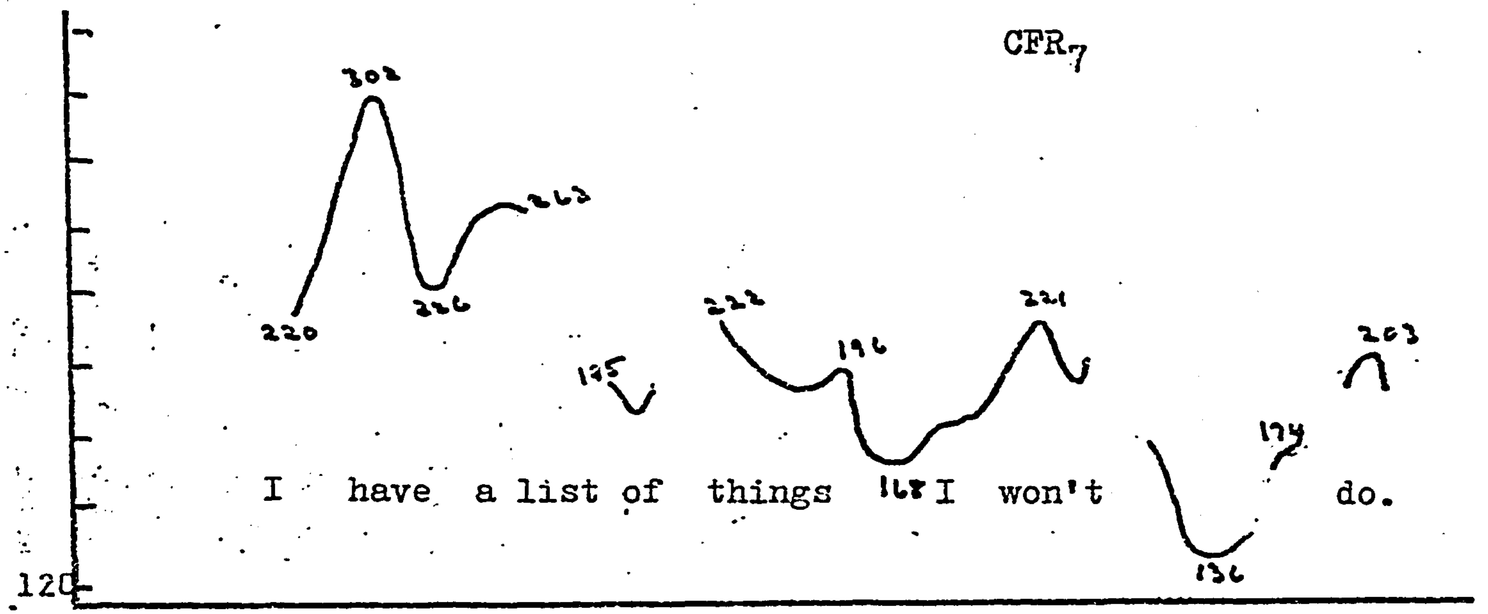

300
60
CF4
269
223
192
202
157
182
184
173
186
169
156
177
203
165
201
170
Everybody's freedom might be considered a threat to someone else

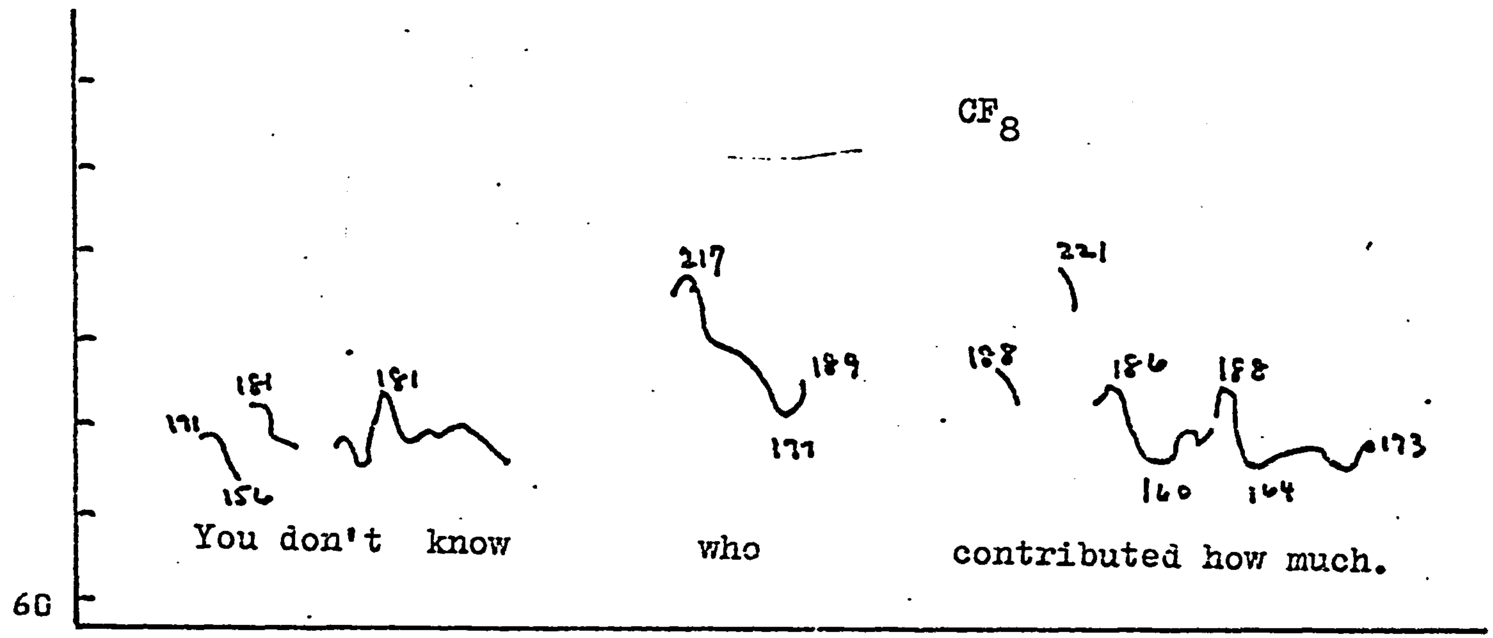

300
60
CFR4
294
203
189
244
180
200
169
197
161
183
156
211
164
201
192
Everybody's freedom might be con sidered a threat to someone else

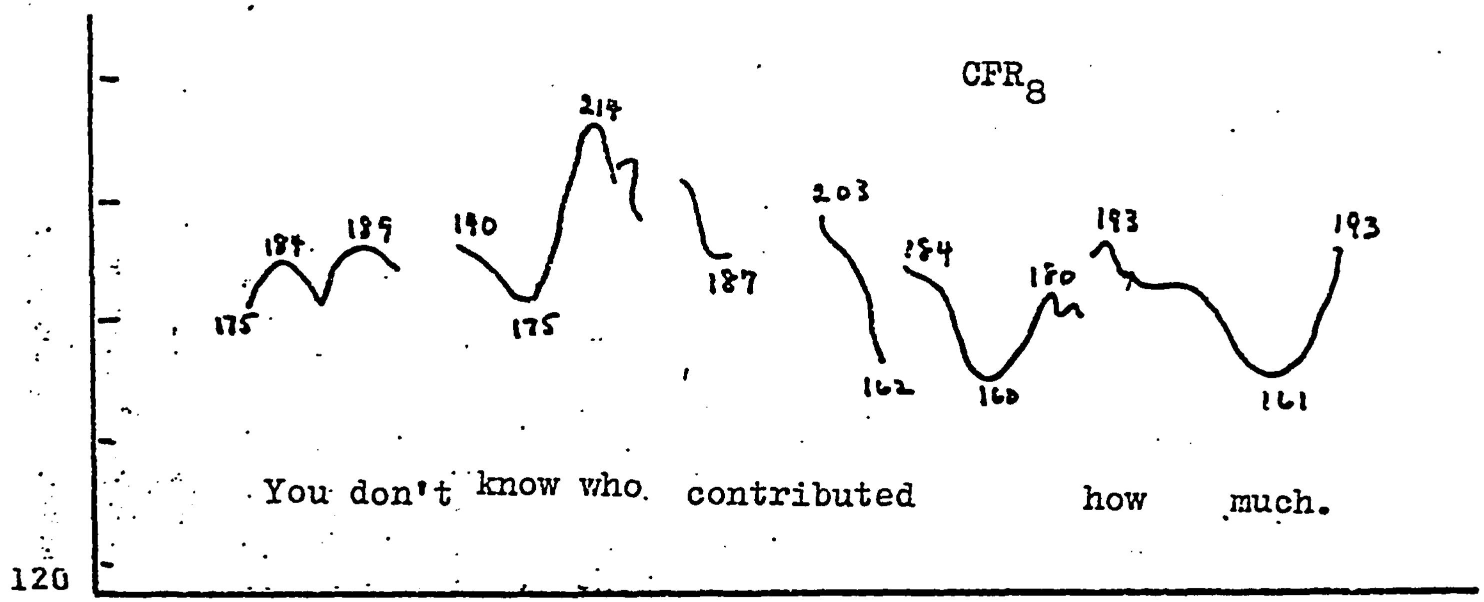

300
60
CF 6
212
179
169
153
162
163
161
I didn't ass ume that

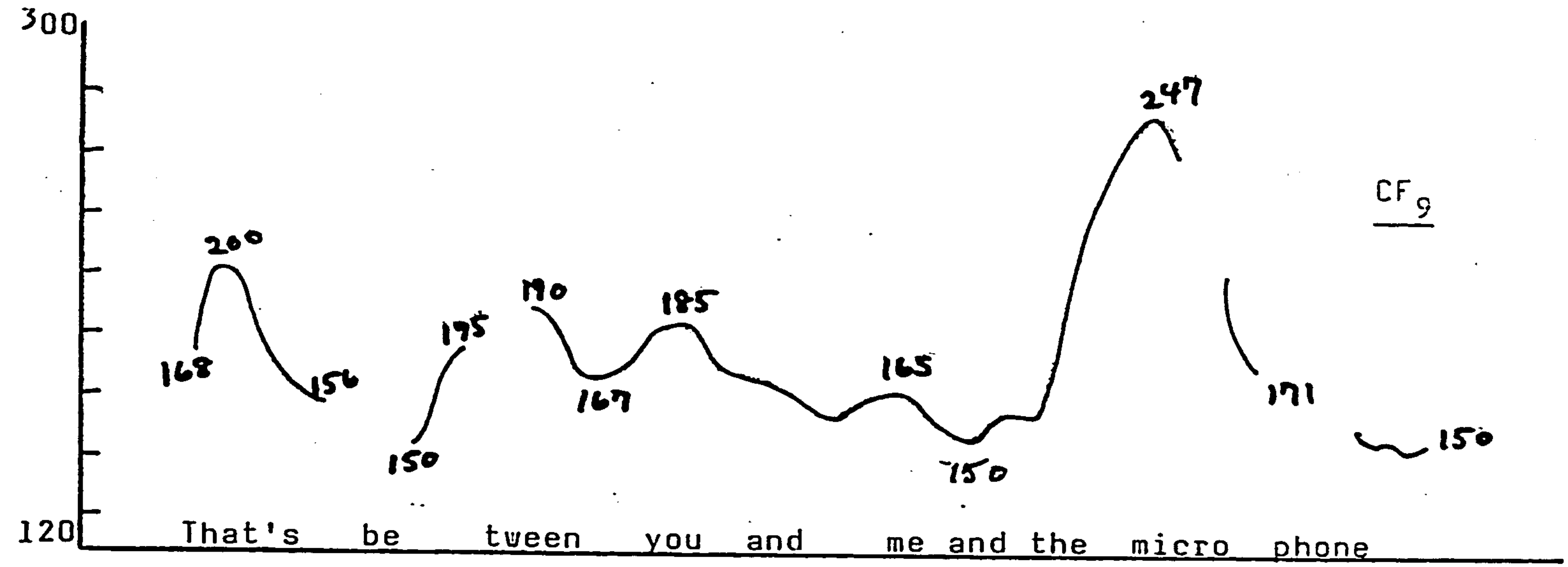

300
60
CFR 6
234
186
172
165
I didn't a ss ume that

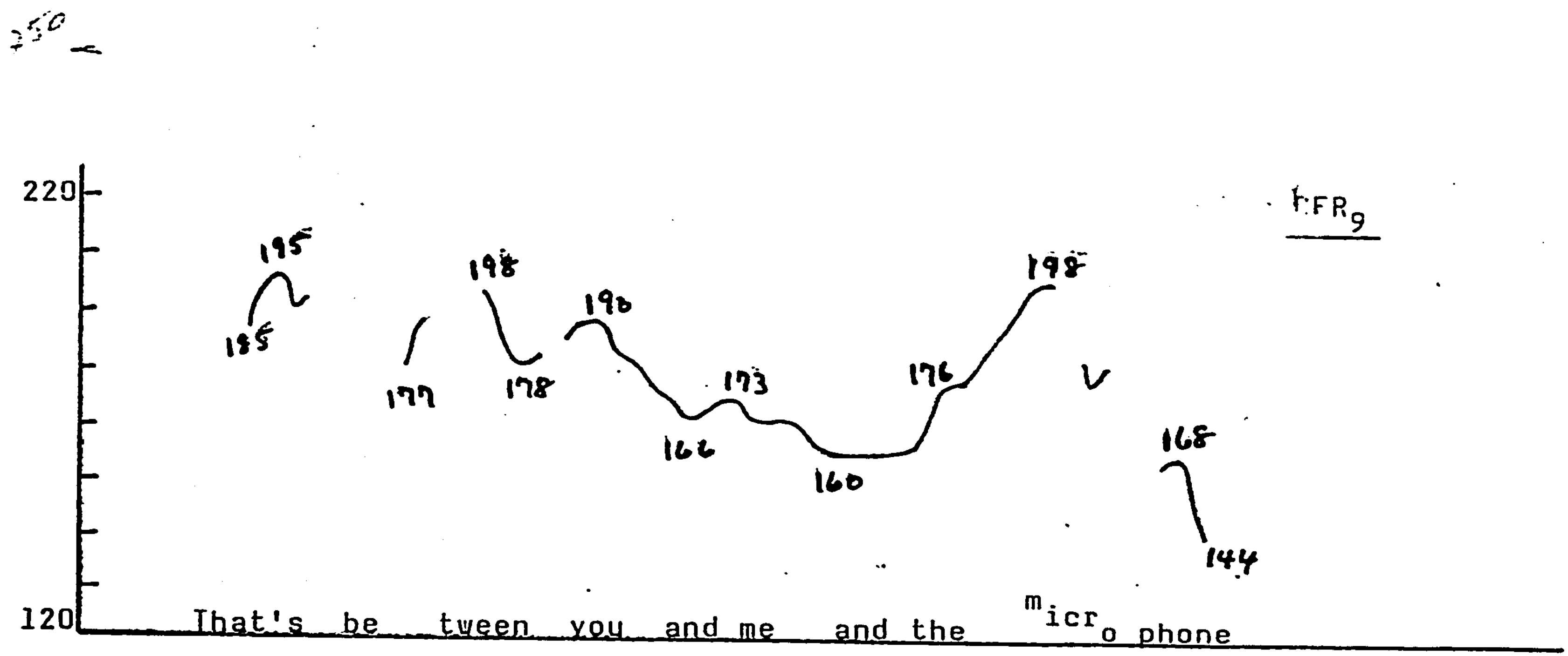

CF7
213
268
206
261
206
186
224
184
199
172
189
208
195
I have a list of things I won't do.
120

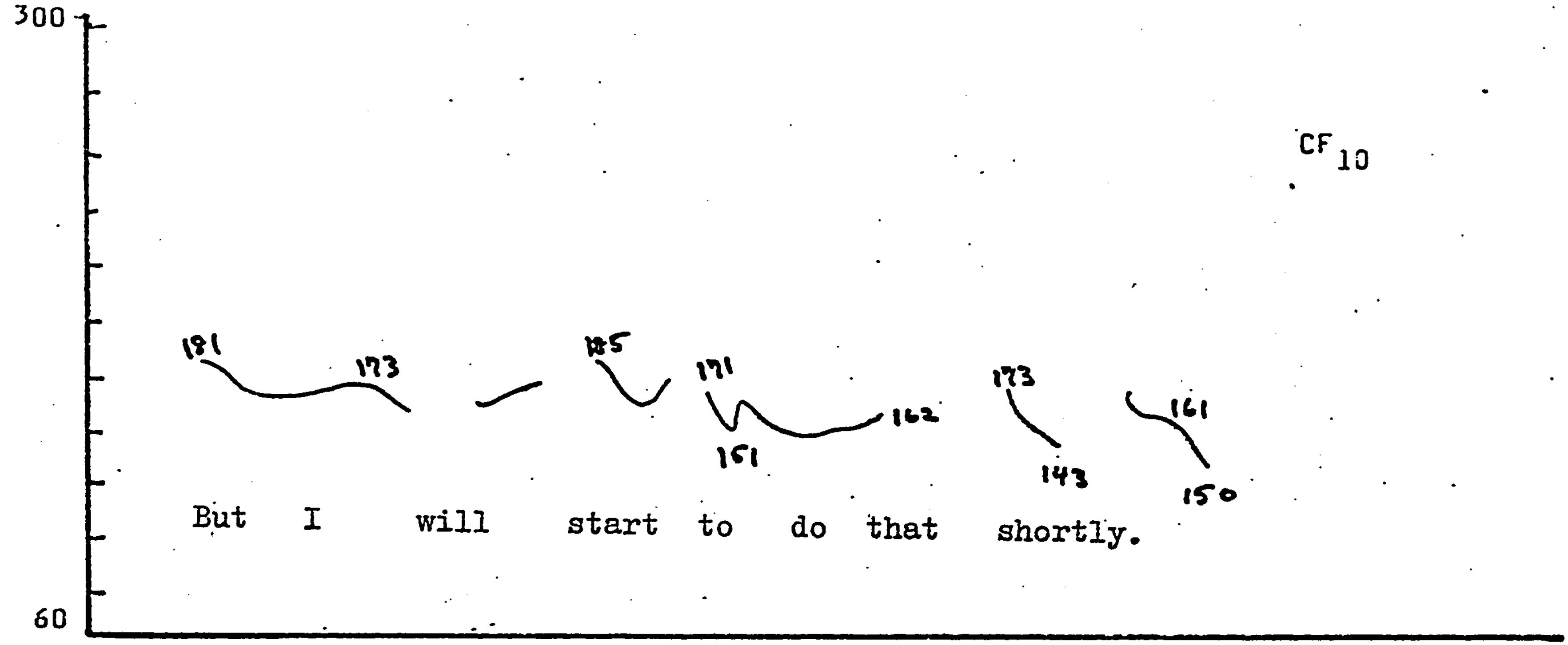

CFR7
302
220
226
262
185
222
196
168
221
136
174
203
I have a list of things I won't do.
120


CF8
217
221
181
181
189
188
186
188
171
156
177
160
164
173
You don't know who contributed how much.
60

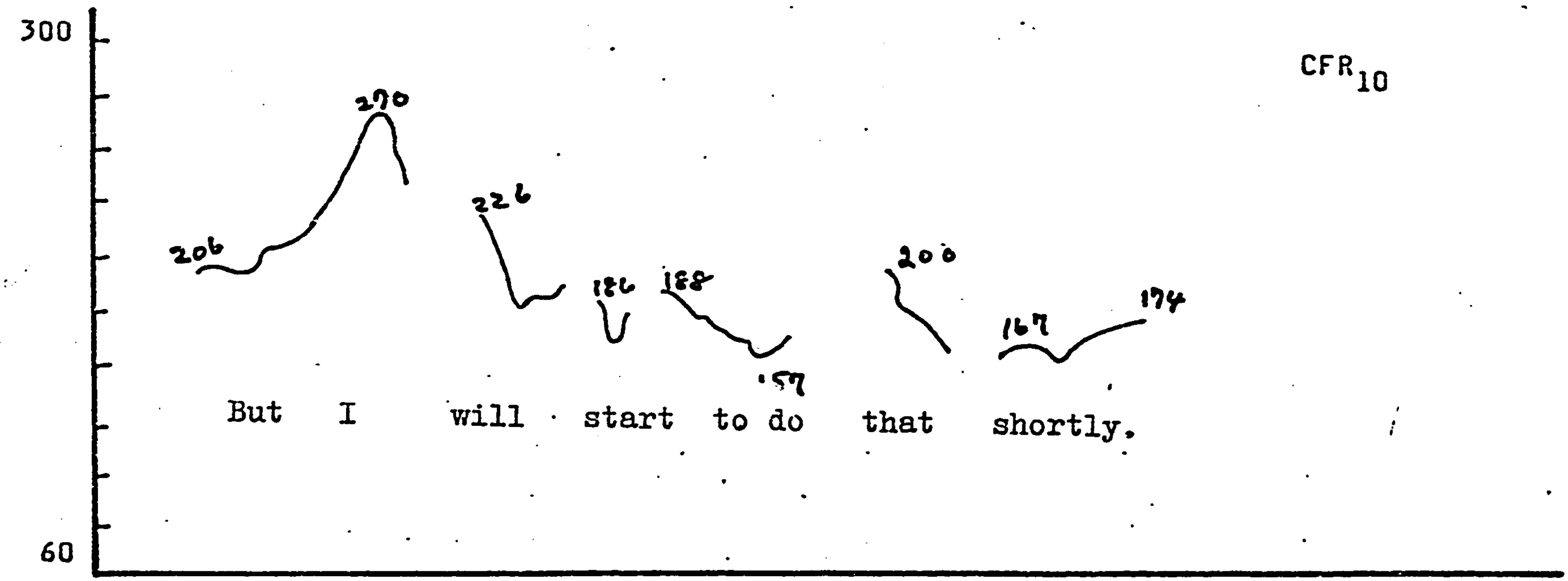

CFR8
214
203
190
184
189
184
193
193
180
175
175
187
162
160
161
You don't know who contributed how much.
120

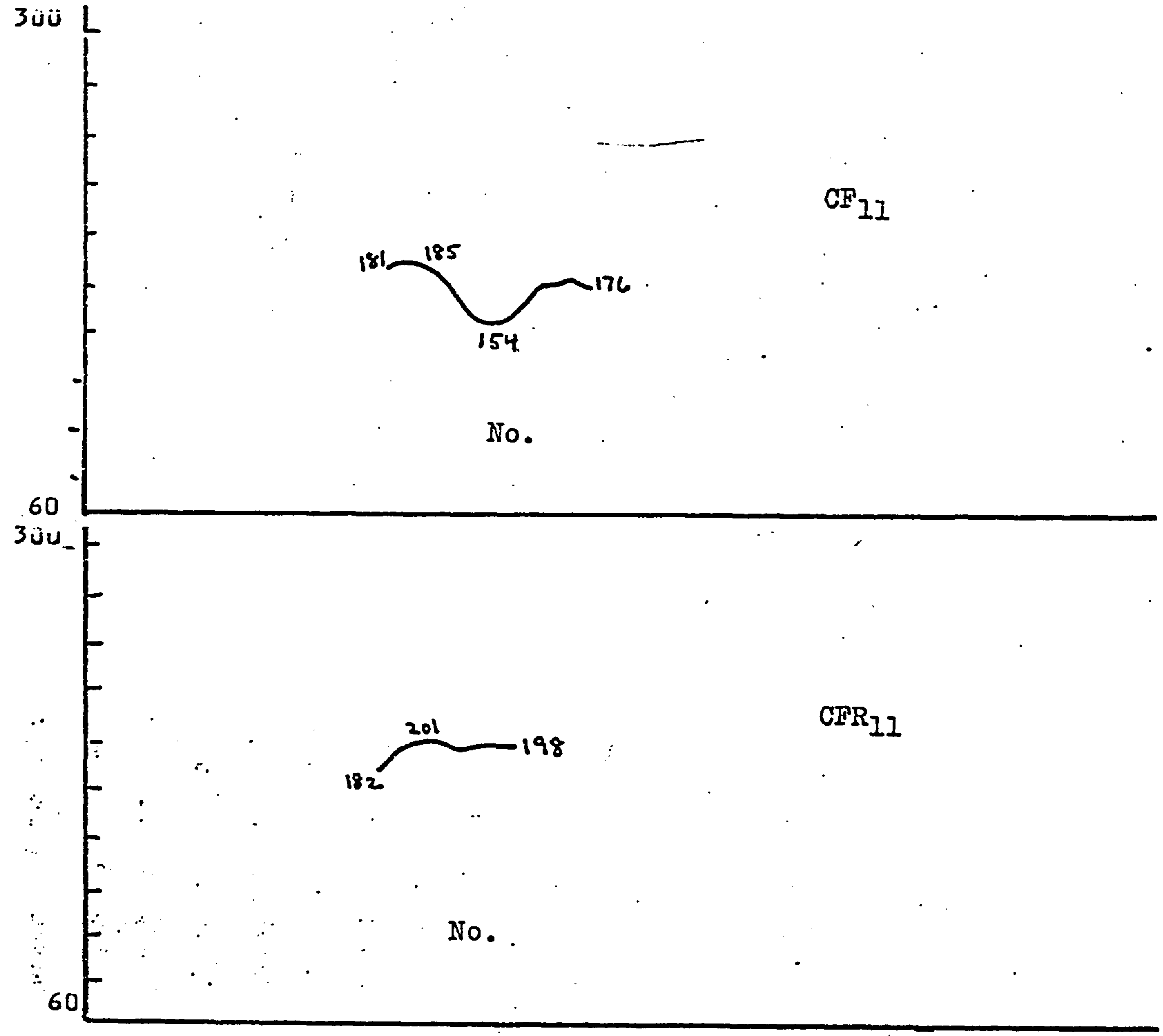

300
120
200
168
156
175
150
190
167
185
165
150
247
171
150
CF_9
That's be tween you and me and the micro phone


250
220
120
195
185
177
198
178
190
166
173
160
176
198
168
144
CFR_9
That's be tween you and me and the micro phone

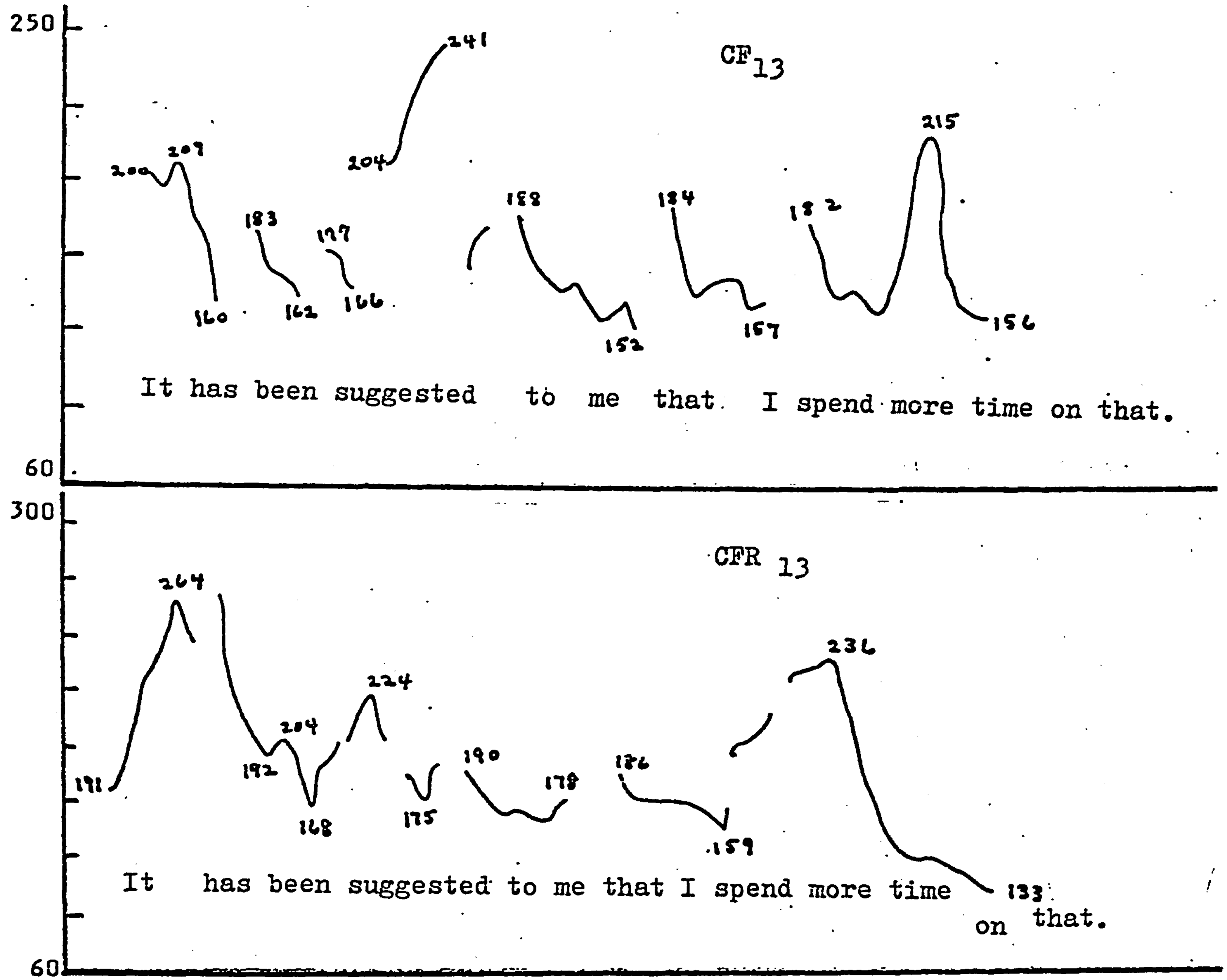


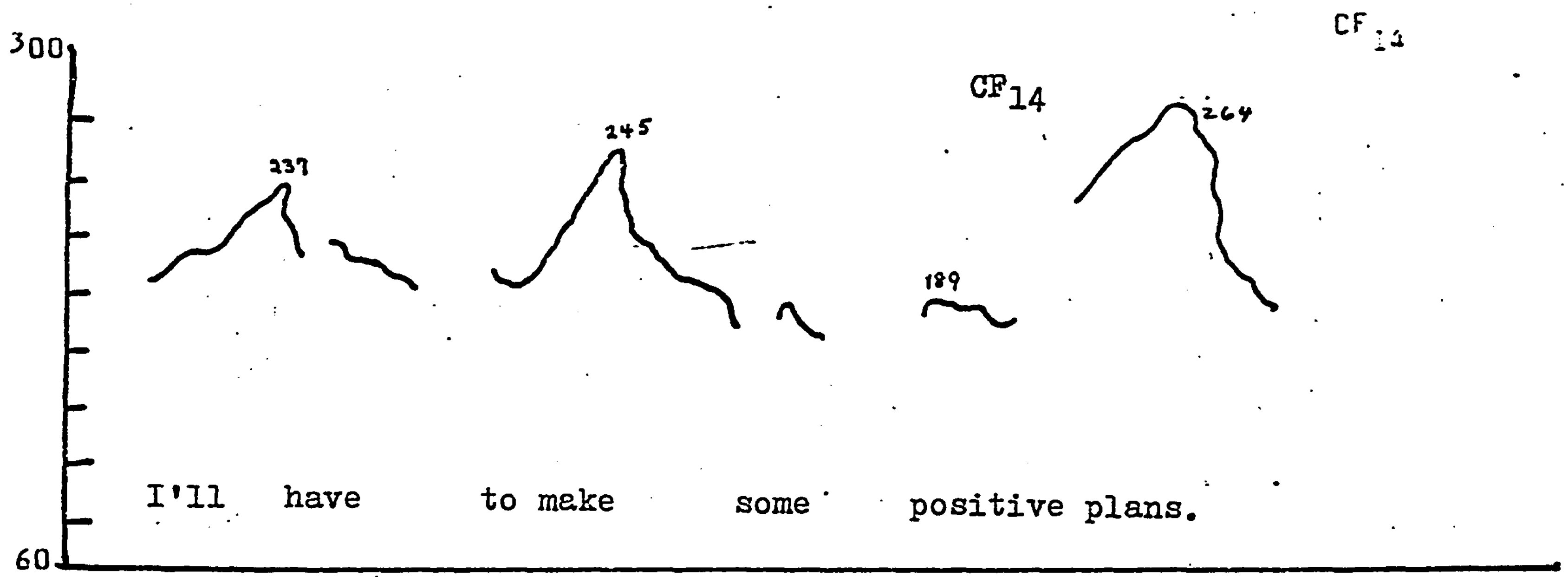
But I will start o

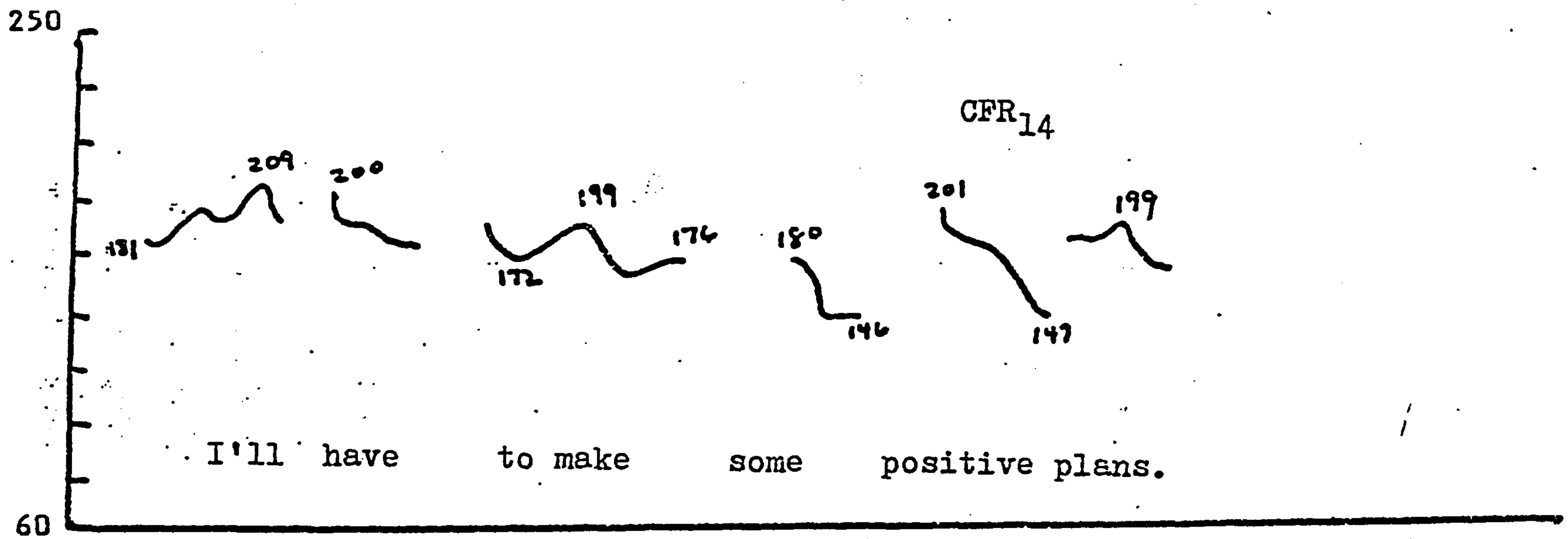


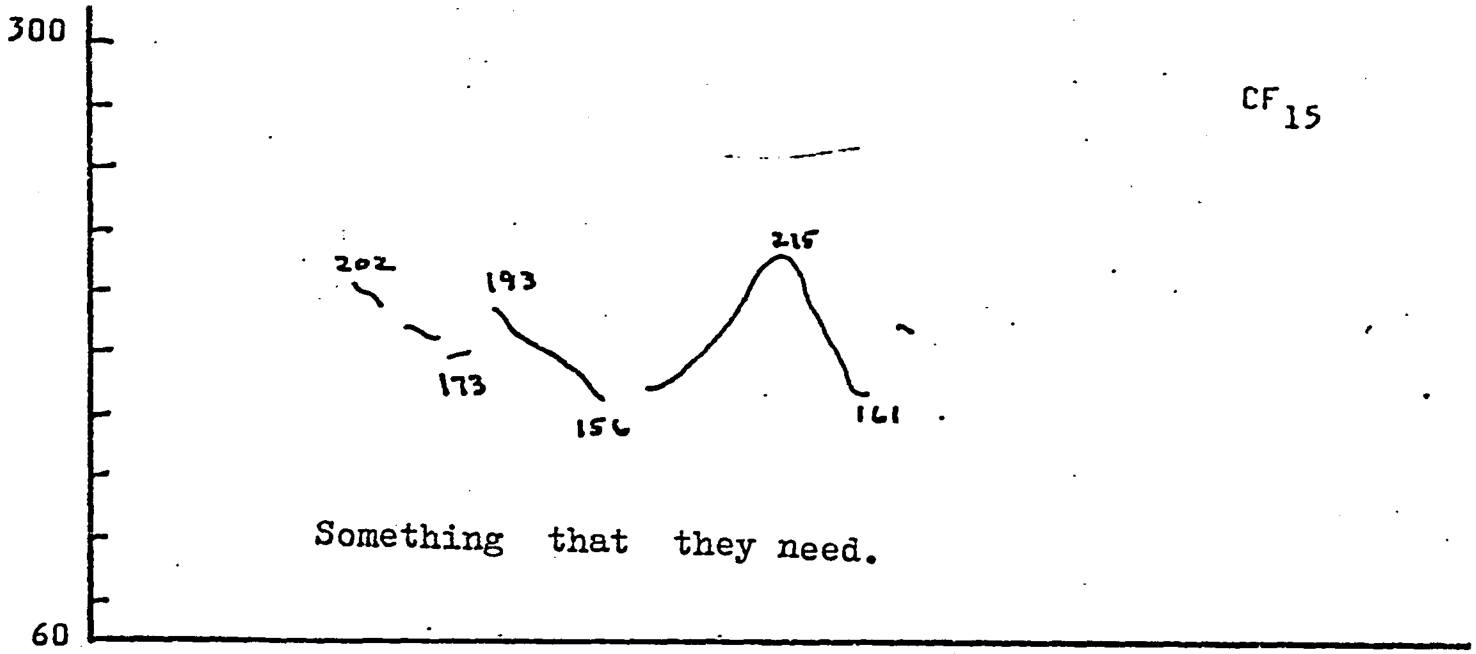


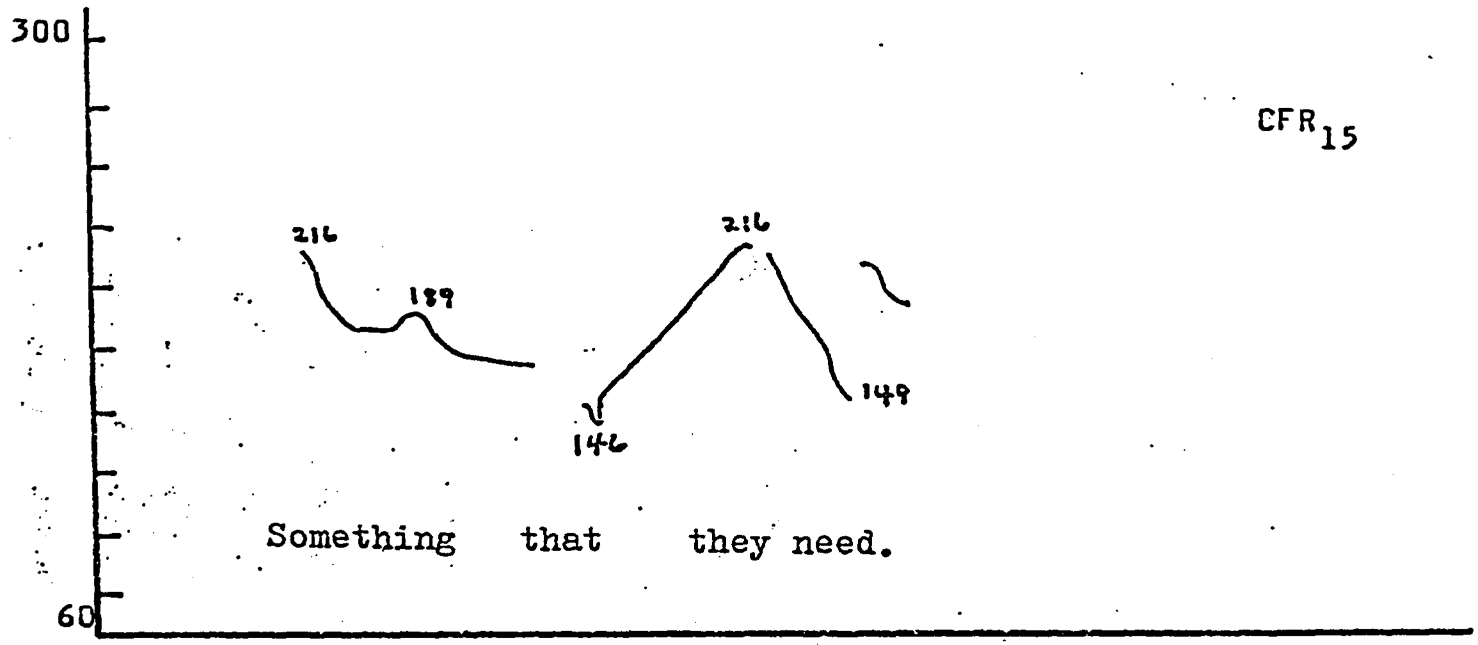
300
60
181
185
154
176
CF$_{11}$
No.
300
60
182
201
198
CFR$_{11}$
No.

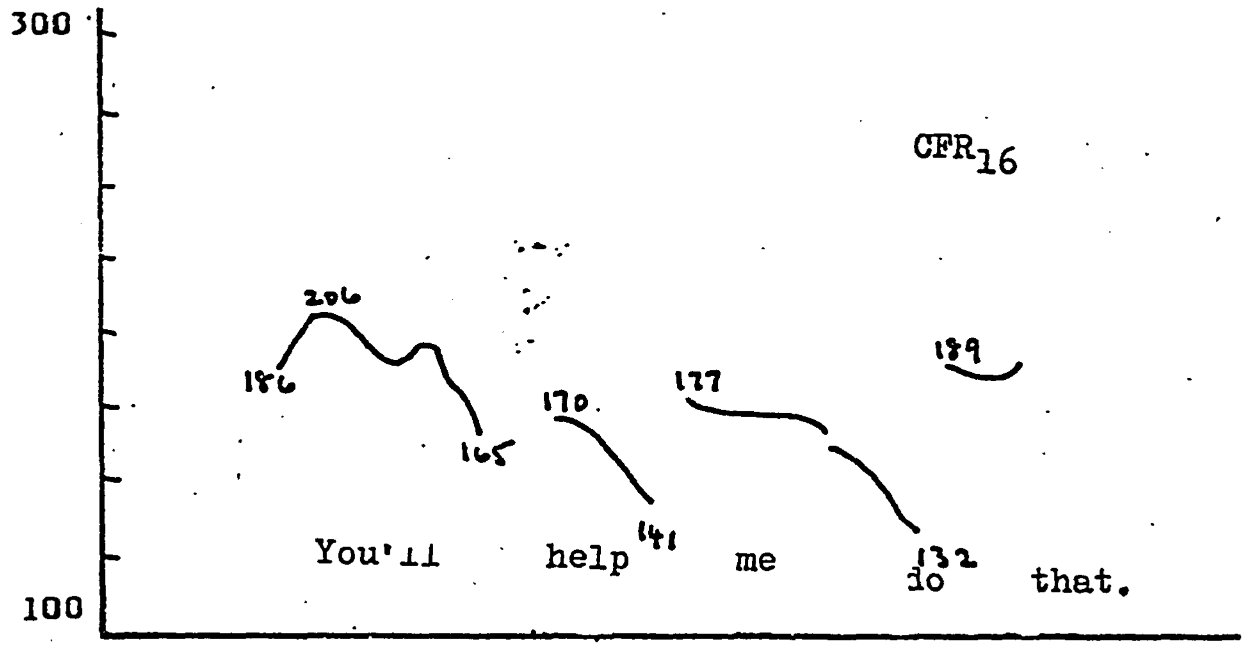

250
60
CF13
241
215
209
200
204
183
177
188
184
182
160
162
166
152
157
156
It has been suggested to me that I spend more time on that.
300
60
CFR 13
264
236
224
204
192
191
168
175
190
178
186
159
133
It has been suggested to me that I spend more time on that.

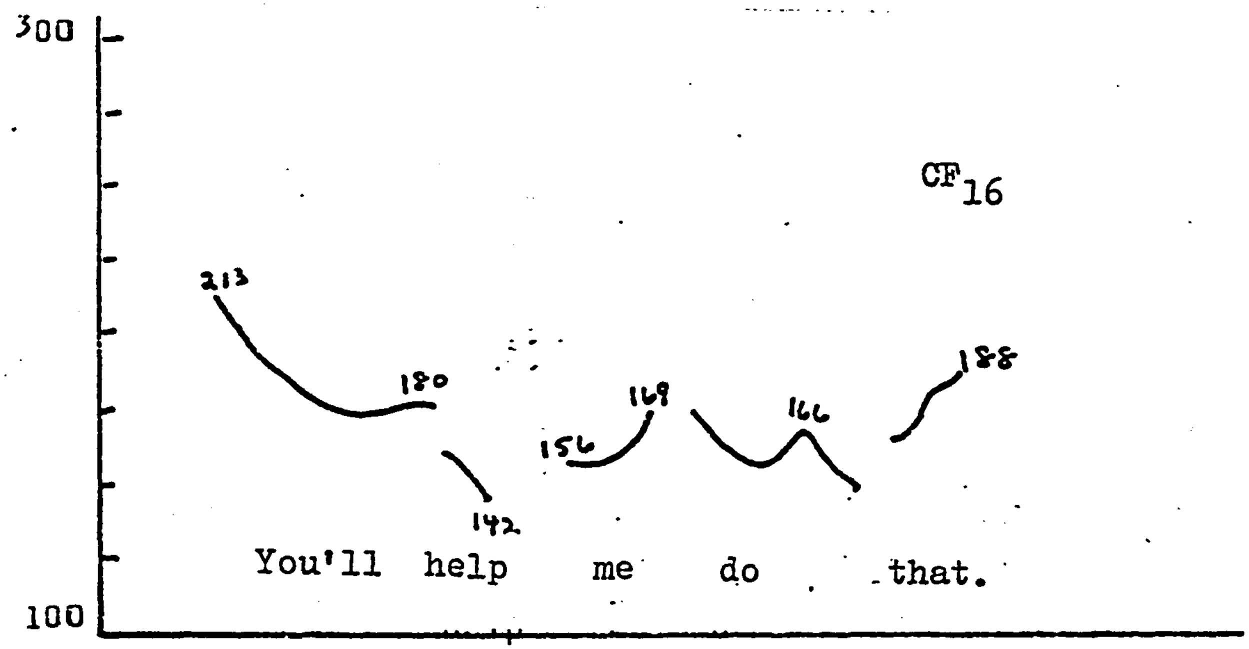

CF14
300
60
237
245
189
264
I'll have to make some positive plans.

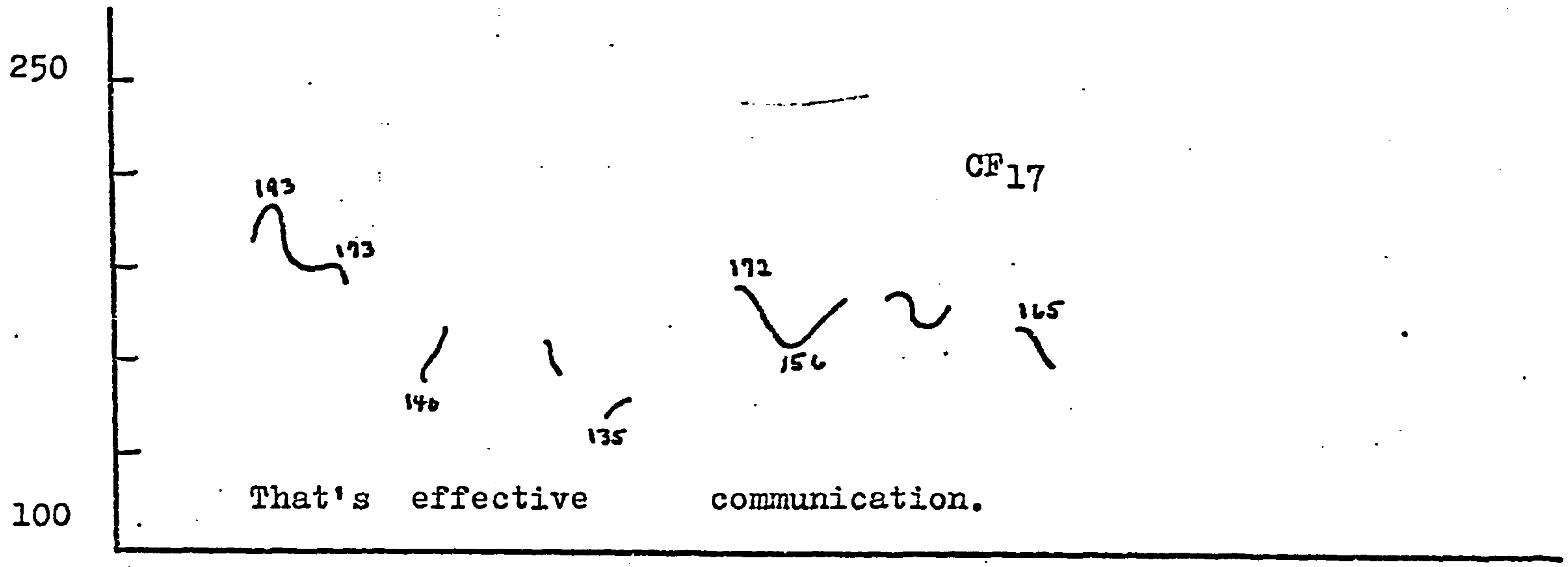

CFR14
250
60
181
209
200
172
199
176
180
146
201
147
199
I'll have to make some positive plans.

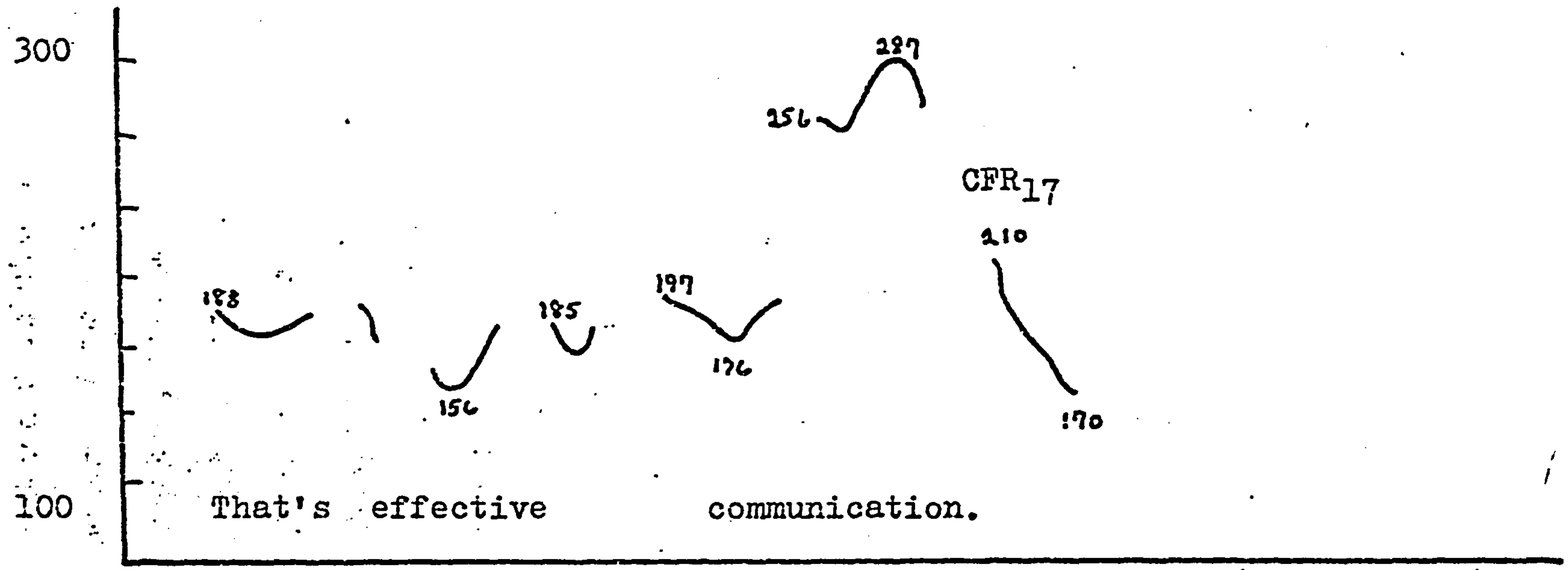

300
60
CF15
202
173
193
156
215
161
Something that they need.

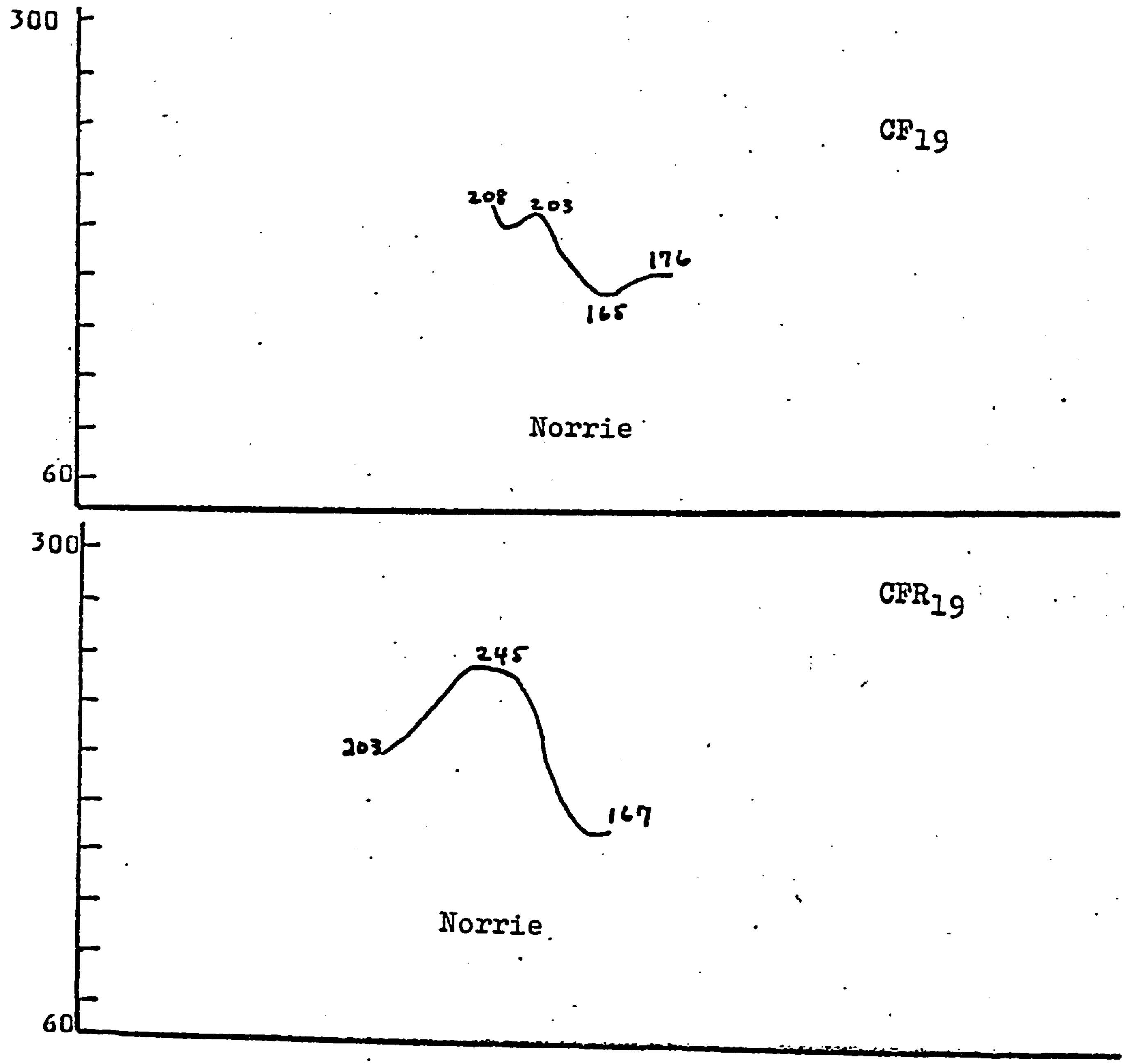

300
60
CFR15
216
189
146
216
149
Something that they need.

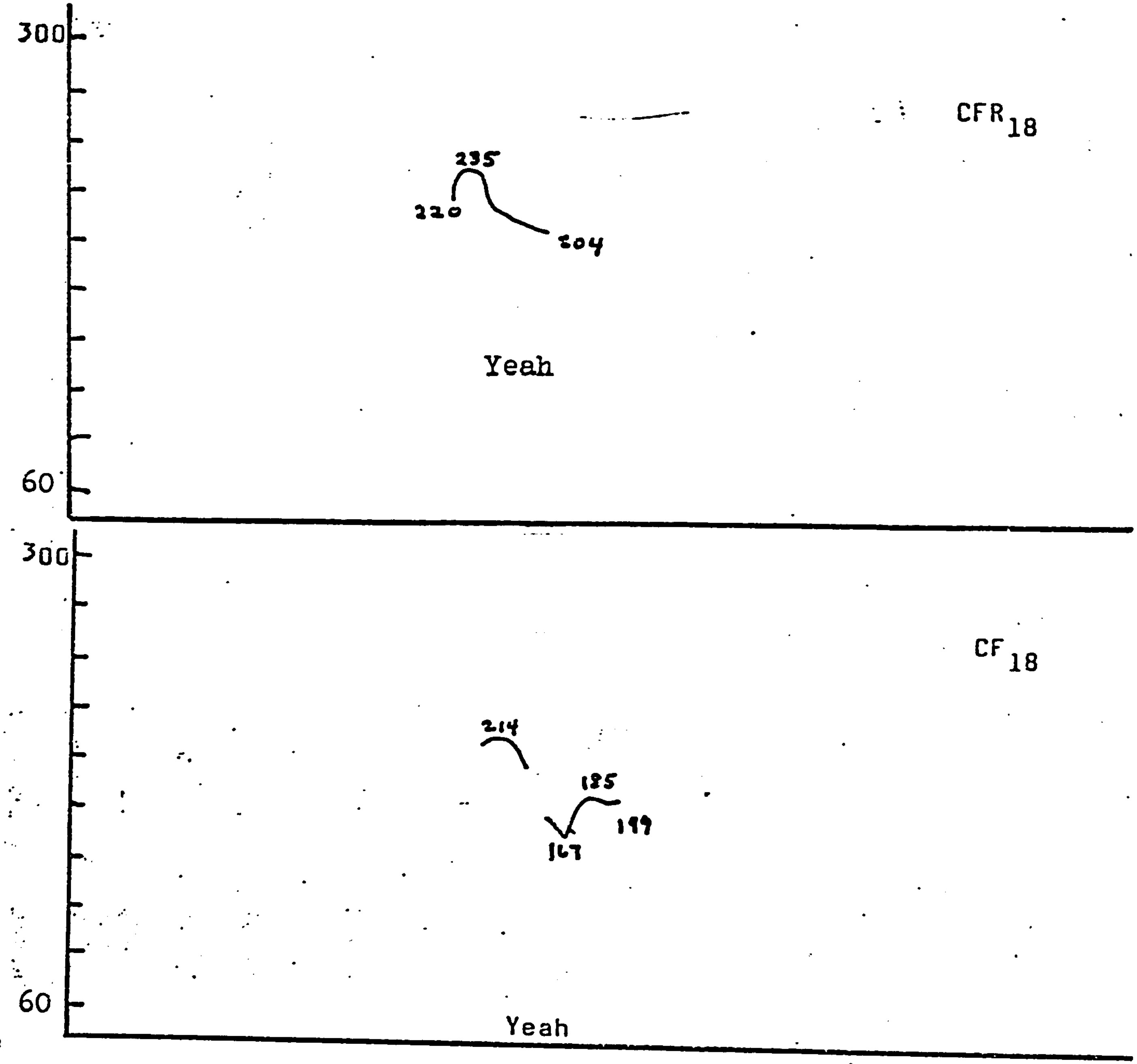

300
100
CFR16
206
186
165
170
141
177
132
189
You'll help me do that.

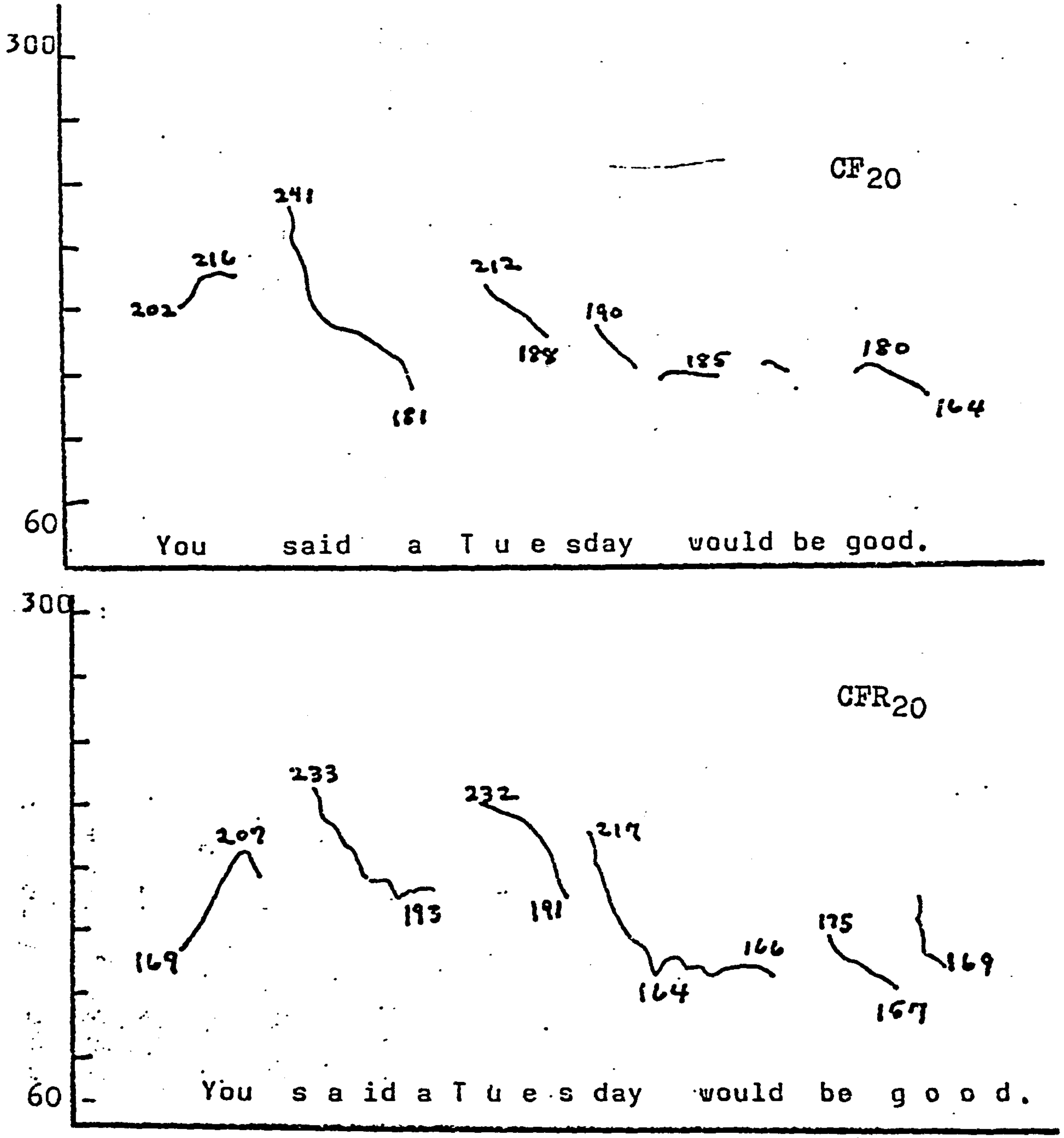

300
100
CF16
213
180
142
156
169
166
188
You'll help me do that.

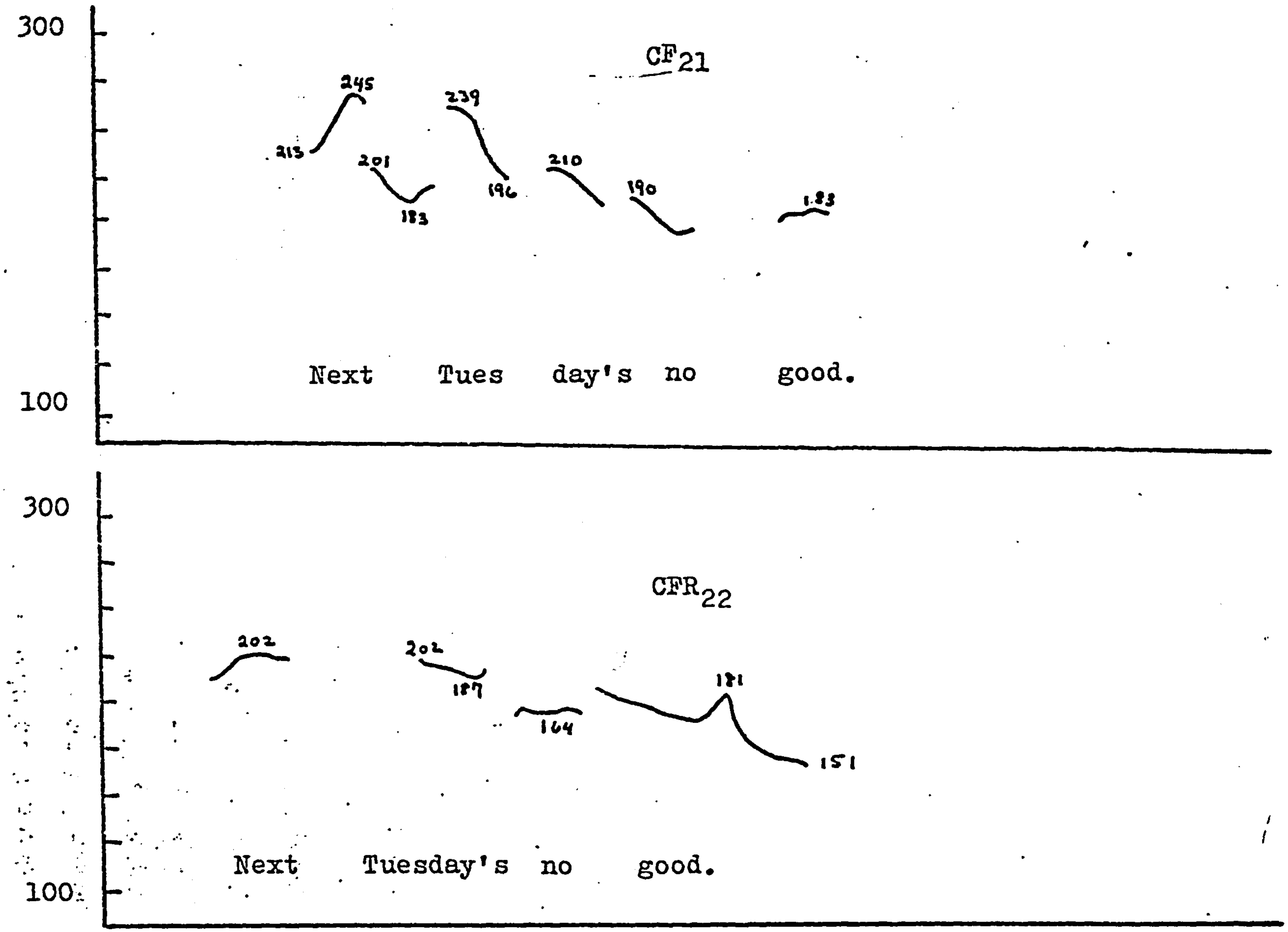

250
100
CF17
193
173
140
135
172
156
165
That's effective communication.

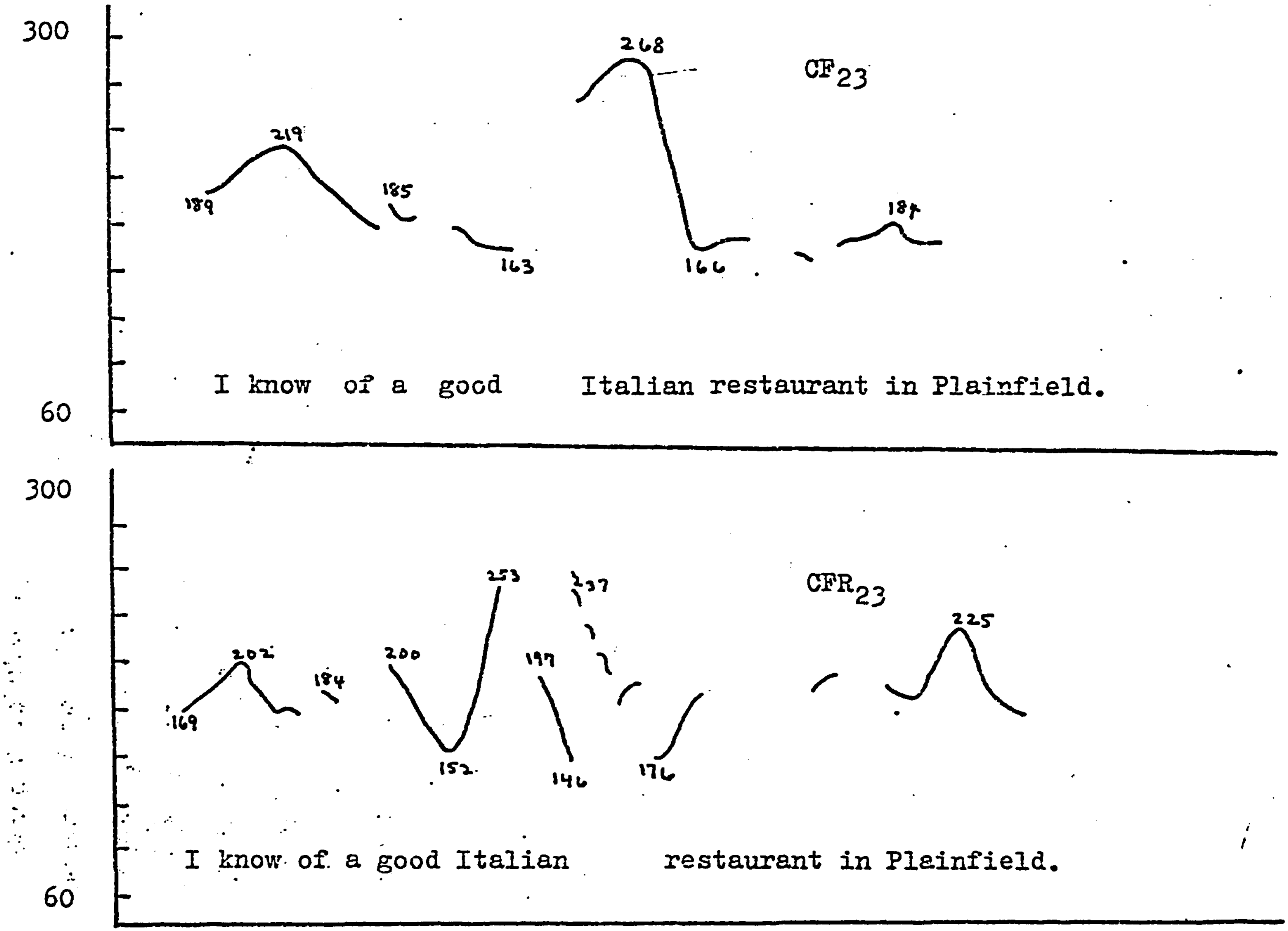

300
100
287
256
CFR17
210
188
156
185
197
176
170
That's effective communication.

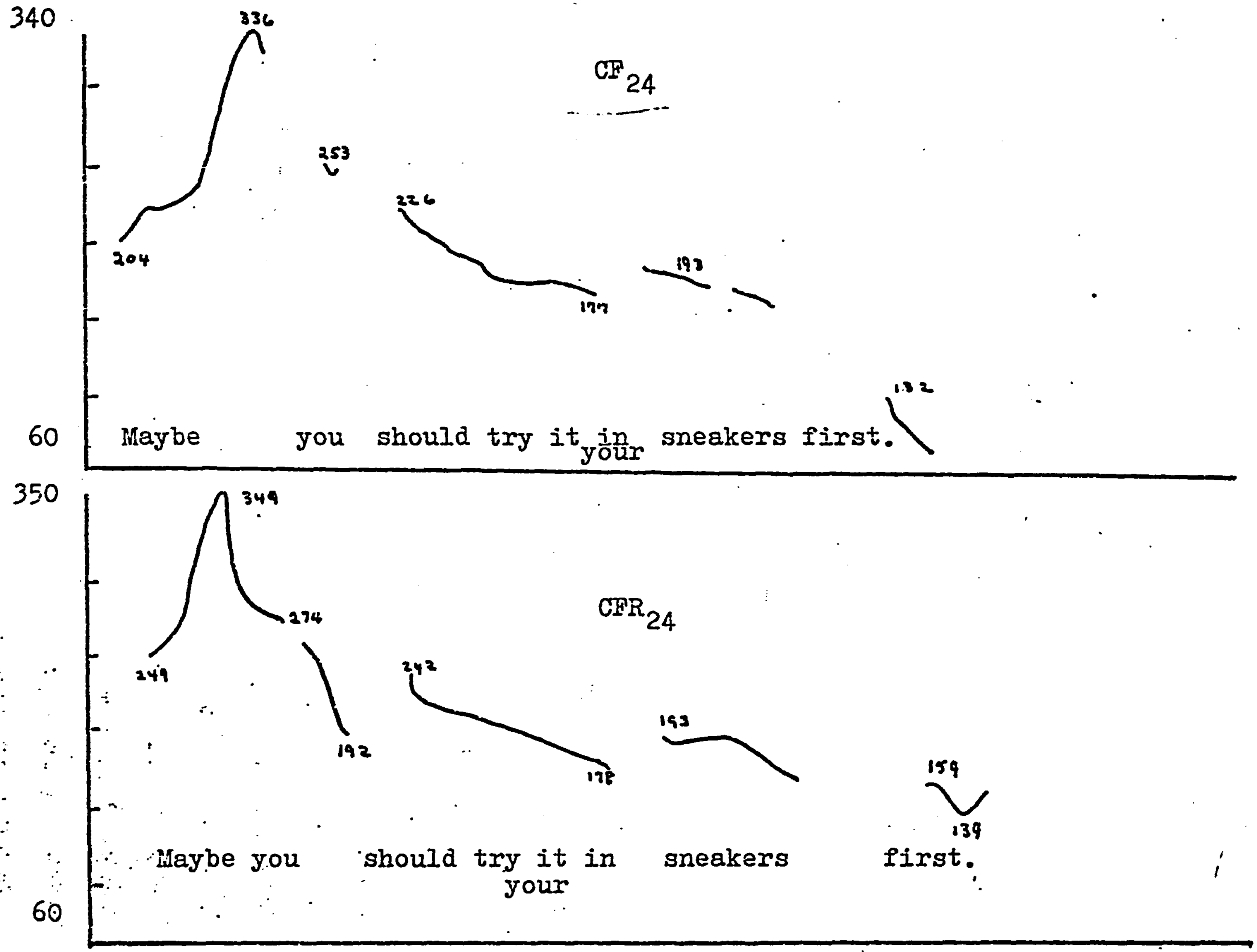

300
60
CF19
208
203
176
165
Norrie
300
60
CFR19
245
203
167
Norrie

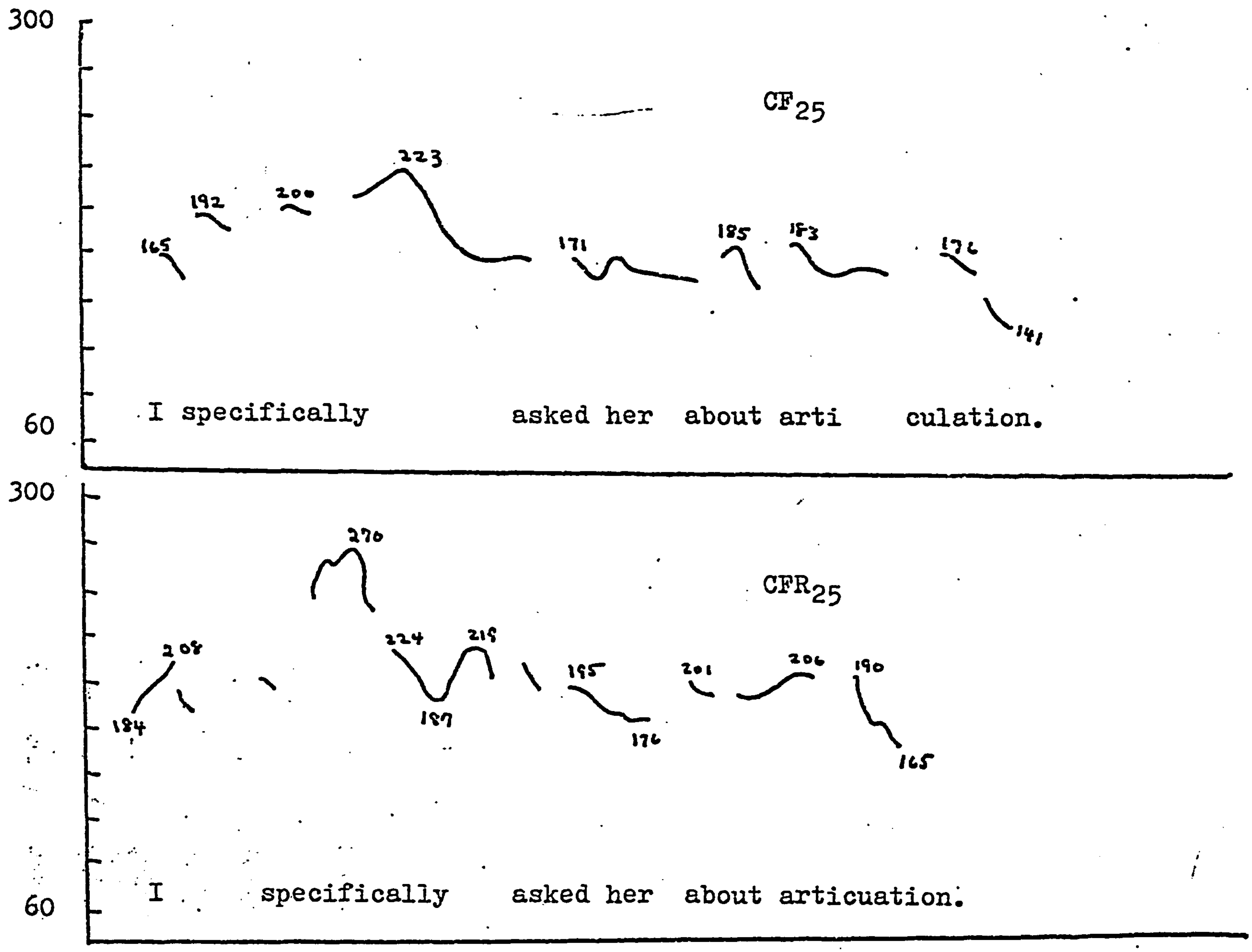

300
60
CFR 18
235
220
204
Yeah
300
60
CF 18
214
185
167
199
Yeah

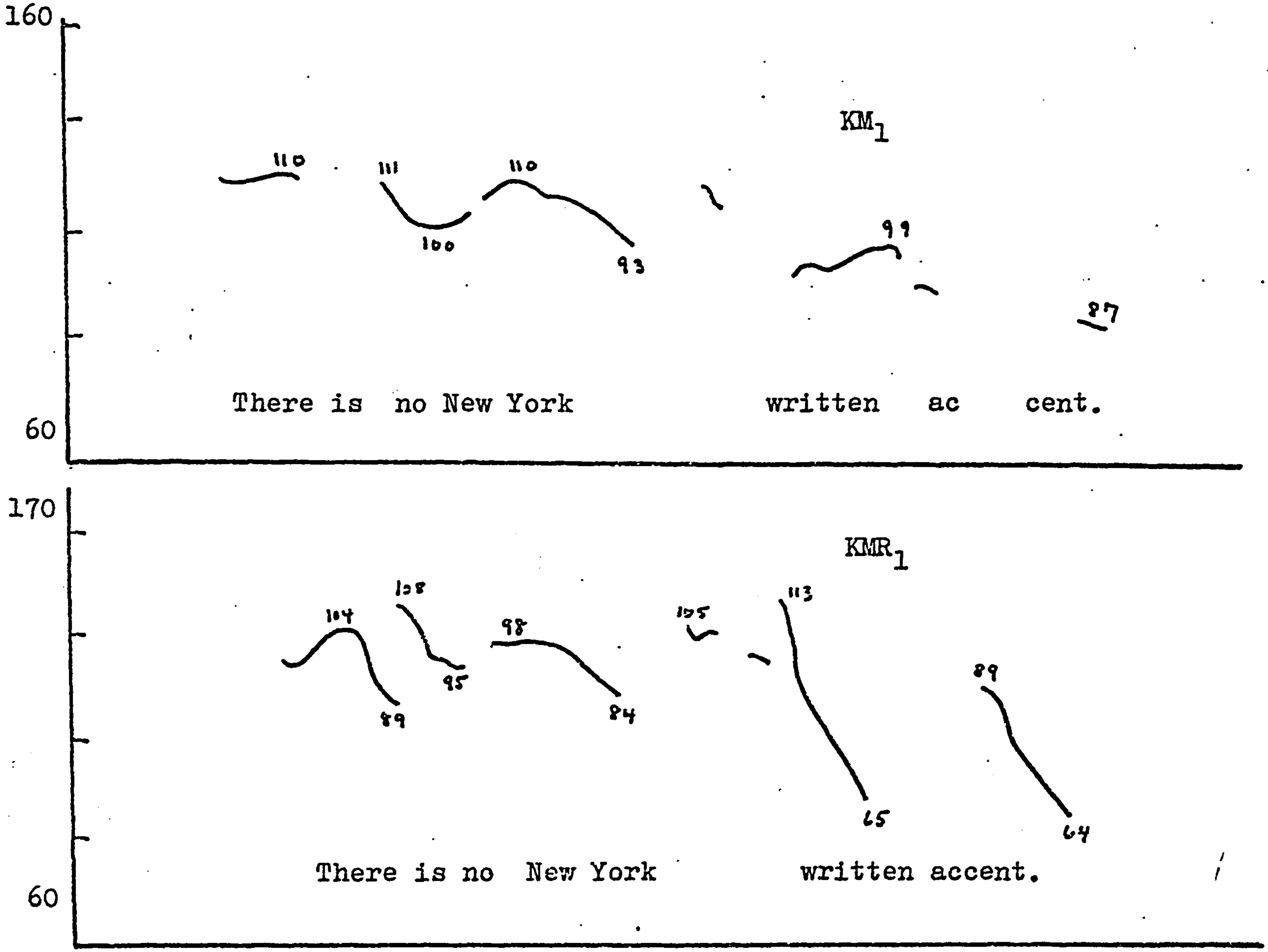

300
60
CF_{20}
202
216
241
181
212
188
190
185
180
164
You said a T u e sday would be good.
300
60
CFR_{20}
169
207
233
193
232
191
217
164
166
175
167
169
You s a id a T u e s day would be g o o d.

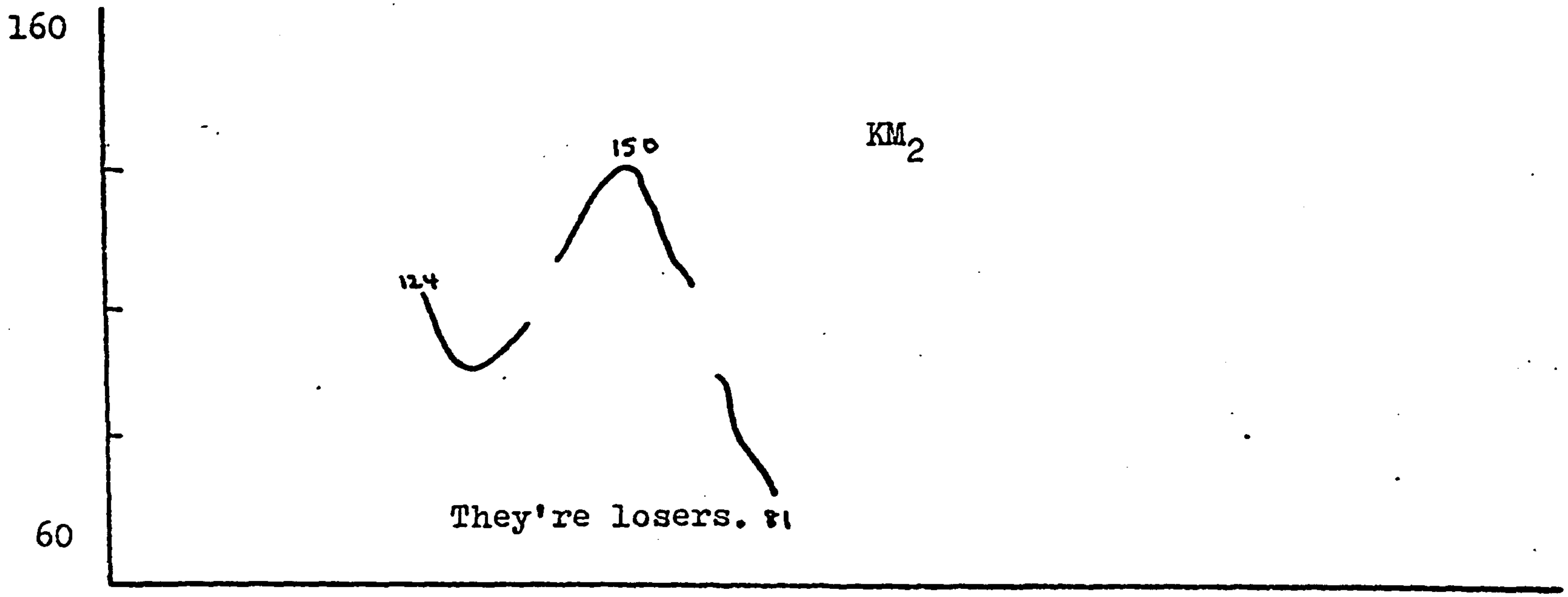

300
100
CF21
213
245
201
183
239
196
210
190
183
Next Tues day's no good.
300
100
CFR22
202
202
187
164
181
151
Next Tuesday's no good.

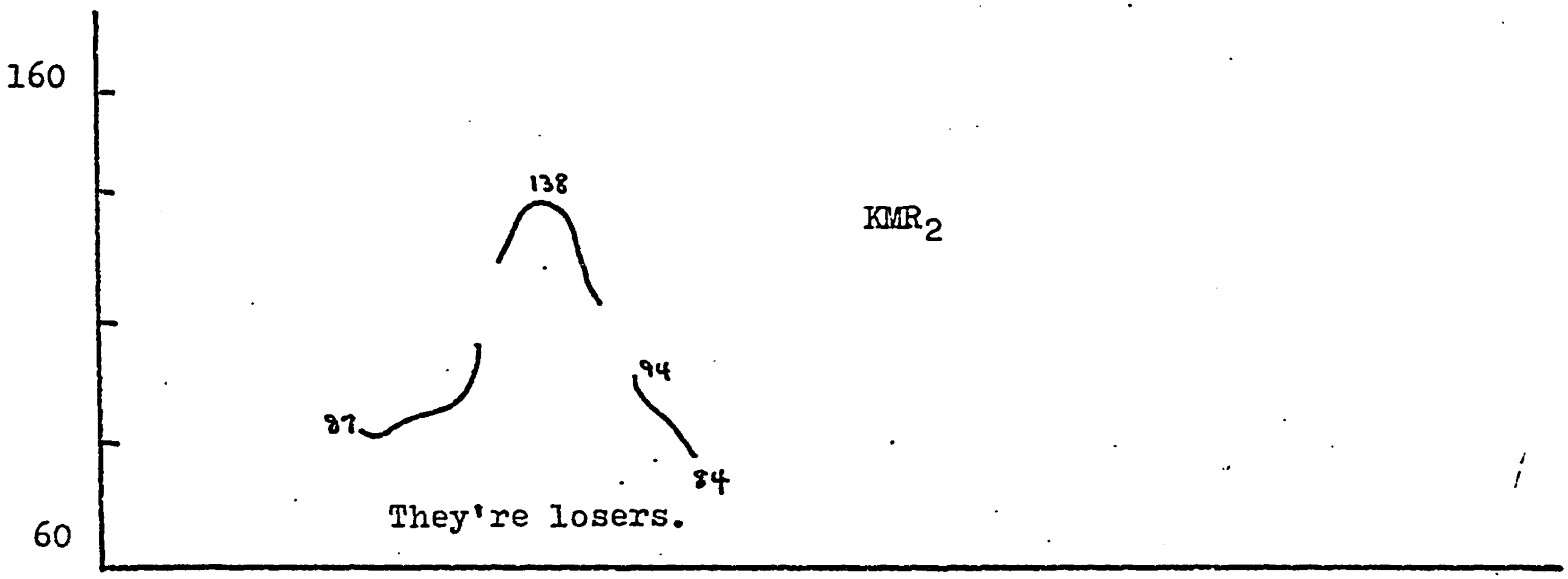

300
268
CF23
219
189
185
184
163
166
I know of a good Italian restaurant in Plainfield.
60
300
253
237
CFR23
225
202
200
197
184
169
152
146
176
I know of a good Italian restaurant in Plainfield.
60

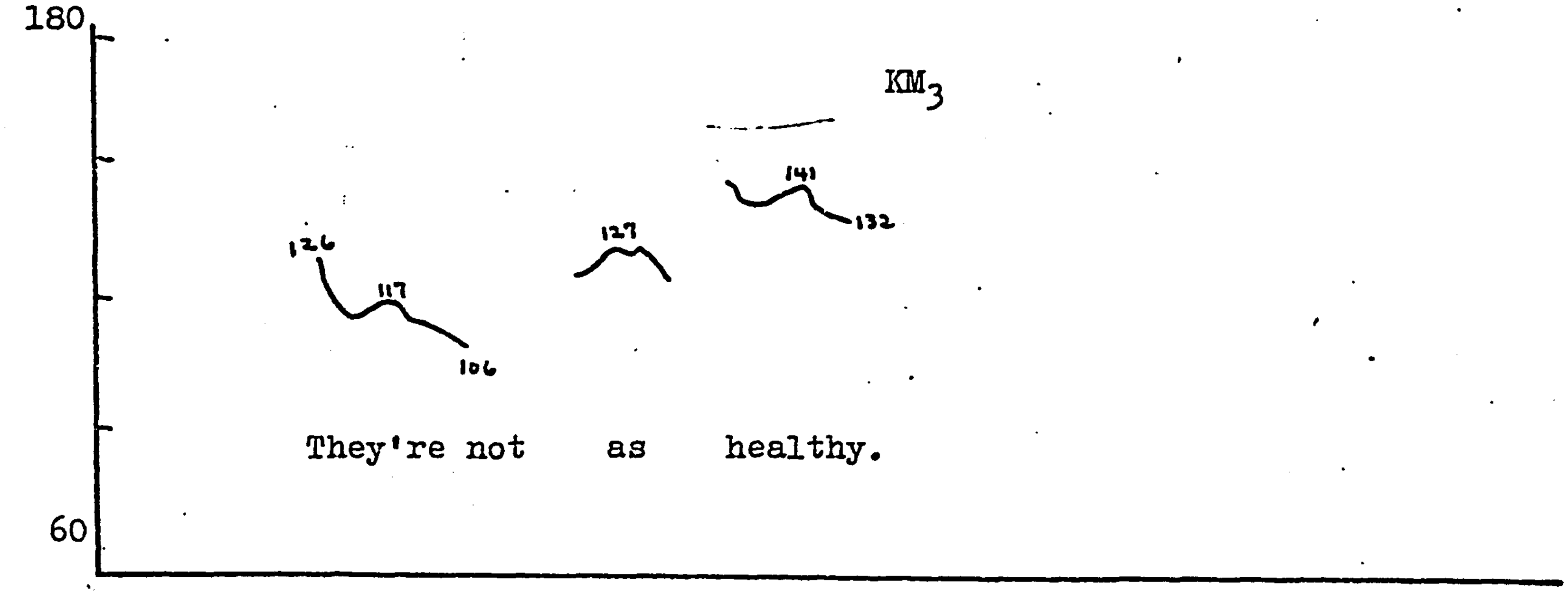

340
336
CF24
253
226
204
193
177
132
60
Maybe you should try it in your sneakers first.
350
349
CFR24
274
249
242
192
193
178
159
139
Maybe you should try it in your sneakers first.
60

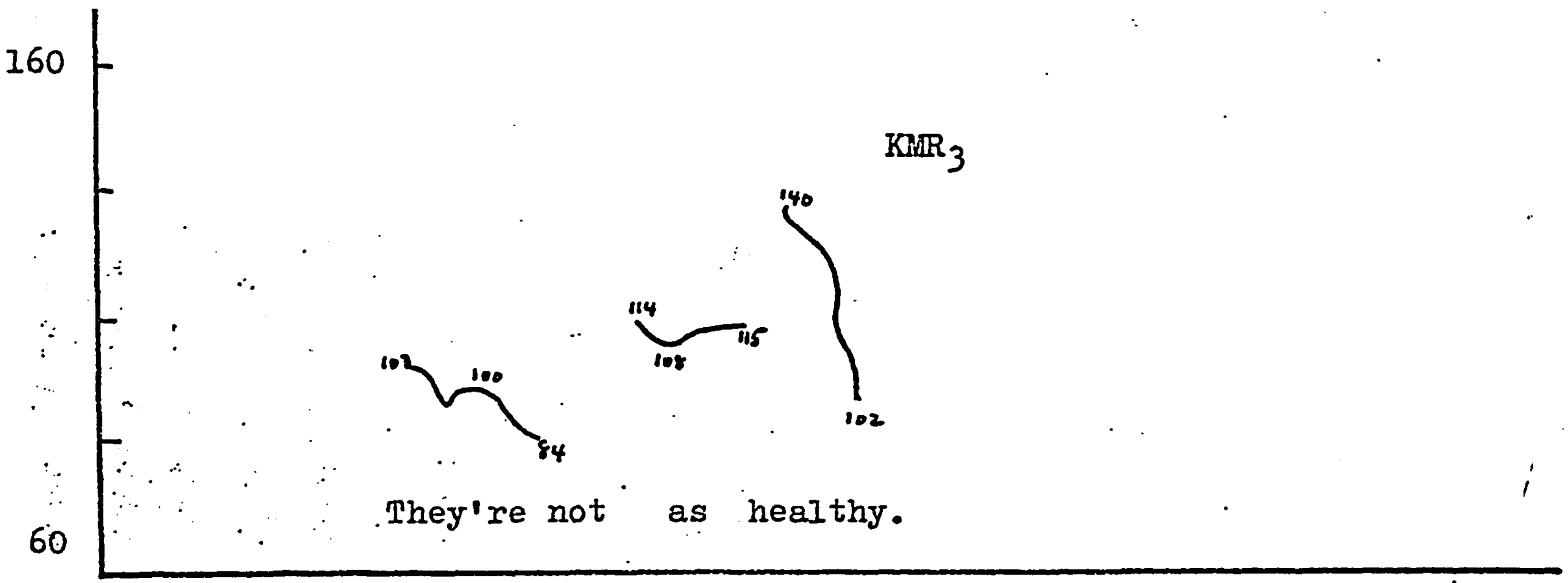

300
CF25
223
192
200
165
171
185
183
176
141
60
I specifically asked her about arti culation.
300
270
CFR25
208
224
219
184
187
195
176
201
206
190
165
60
I specifically asked her about articuation.

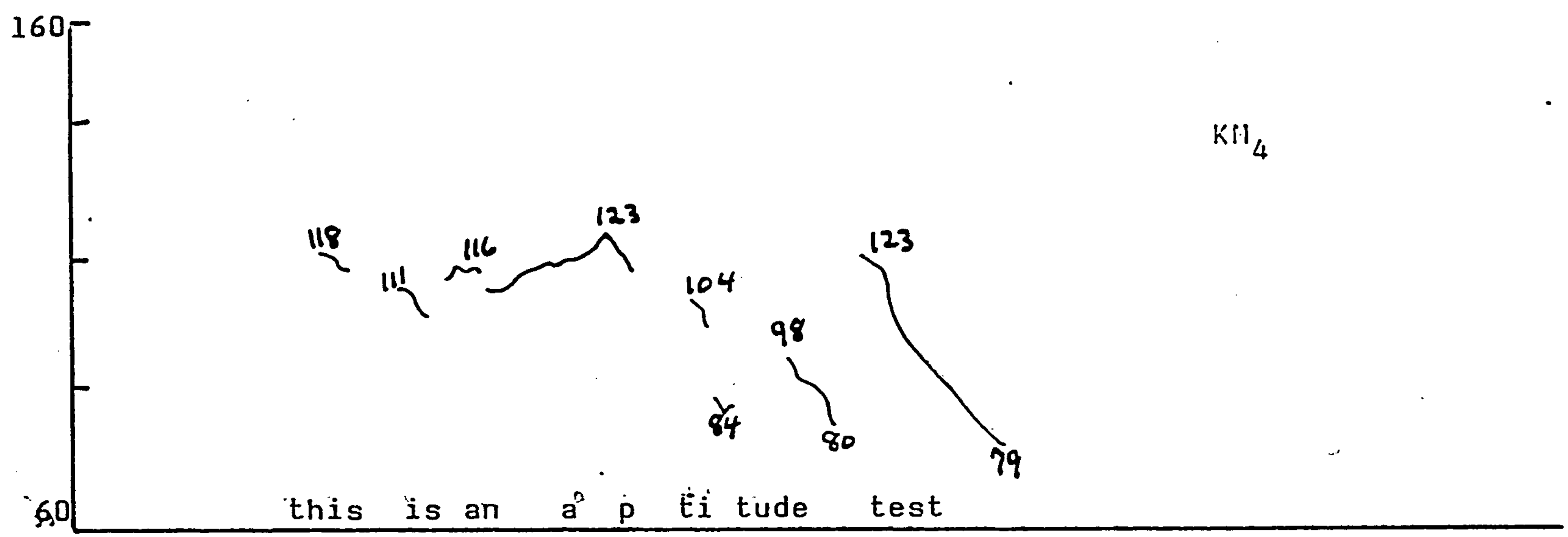

160
KM1
110
111
100
110
93
99
87
There is no New York
written
ac
cent.
60
170
KMR1
104
108
89
95
98
84
105
113
65
89
64
There is no New York
written accent.
60

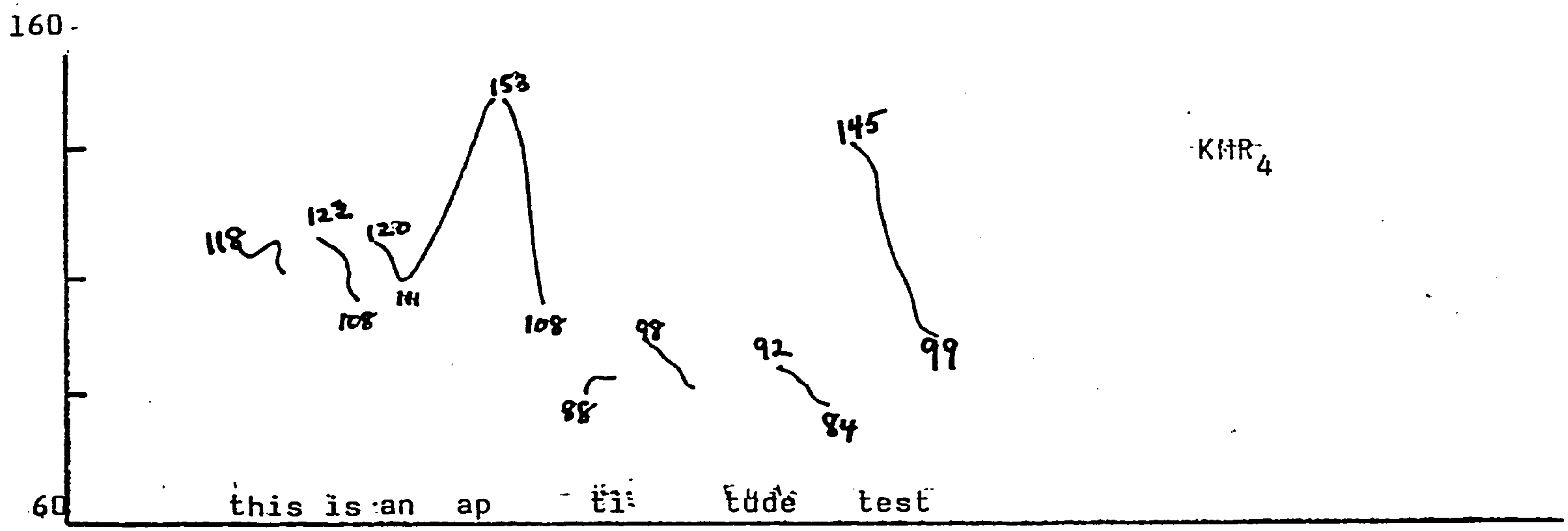

160
60
150
124
81
KM2
They're losers.

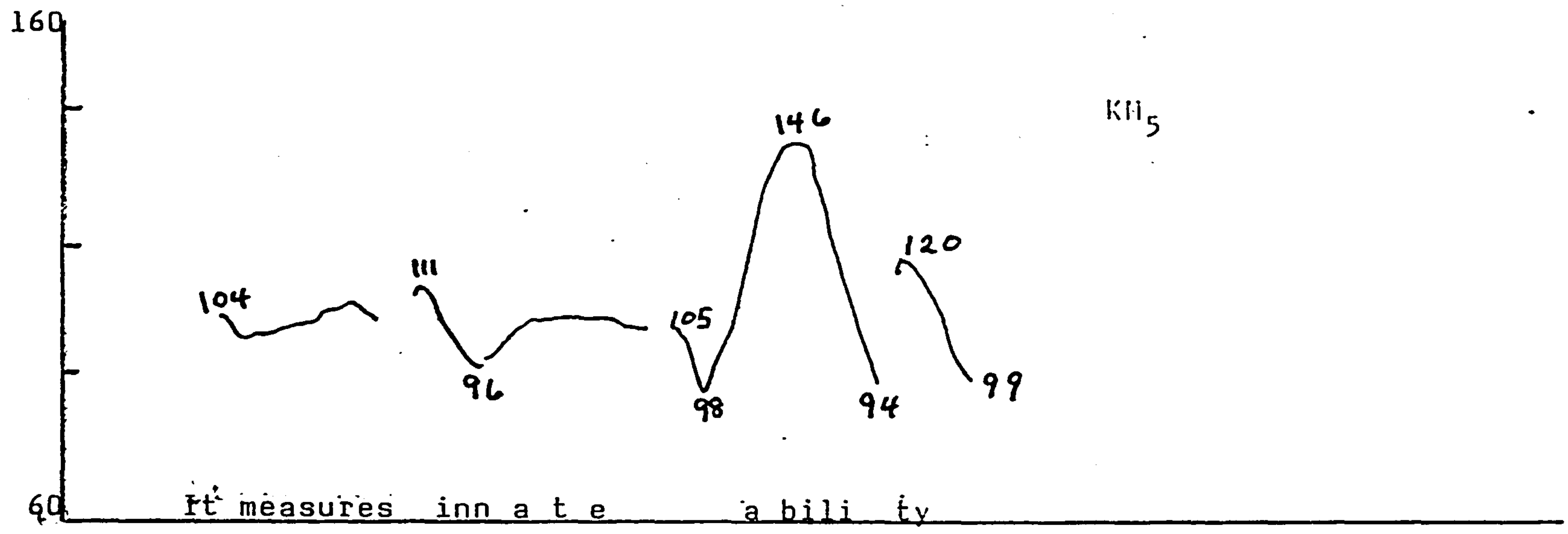

160
60
138
87
94
84
KMR2
They're losers.

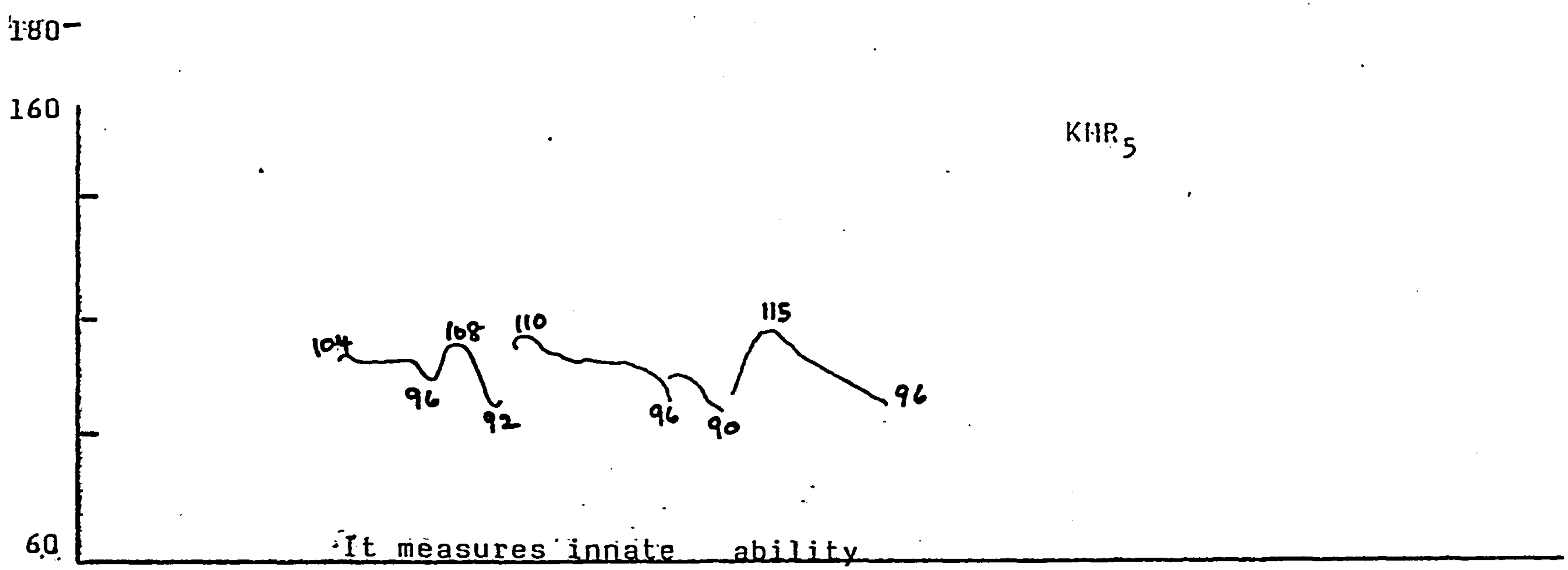

180
60
KM3
They're not as healthy.

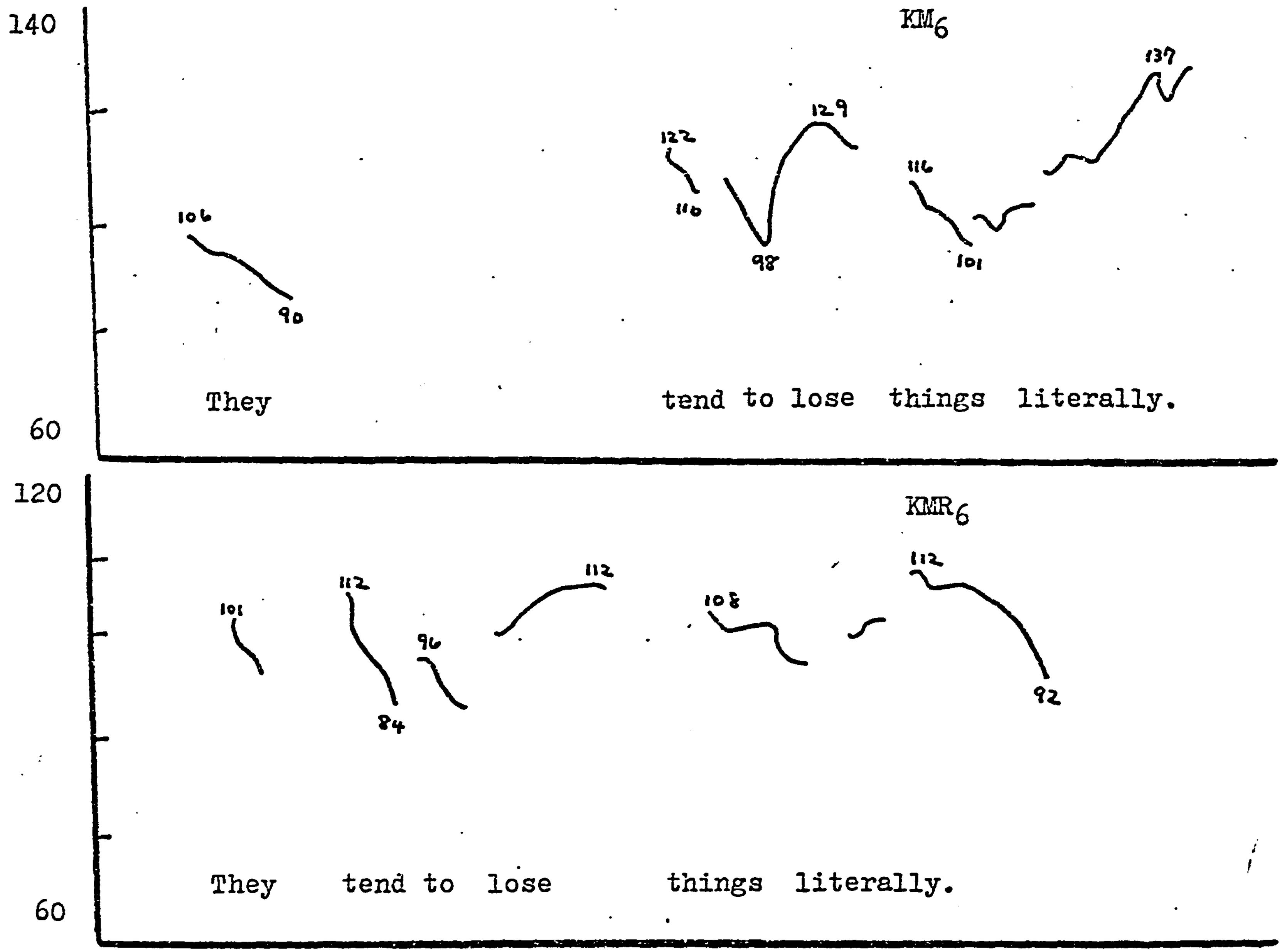

160
60
KMR3
They're not as healthy.

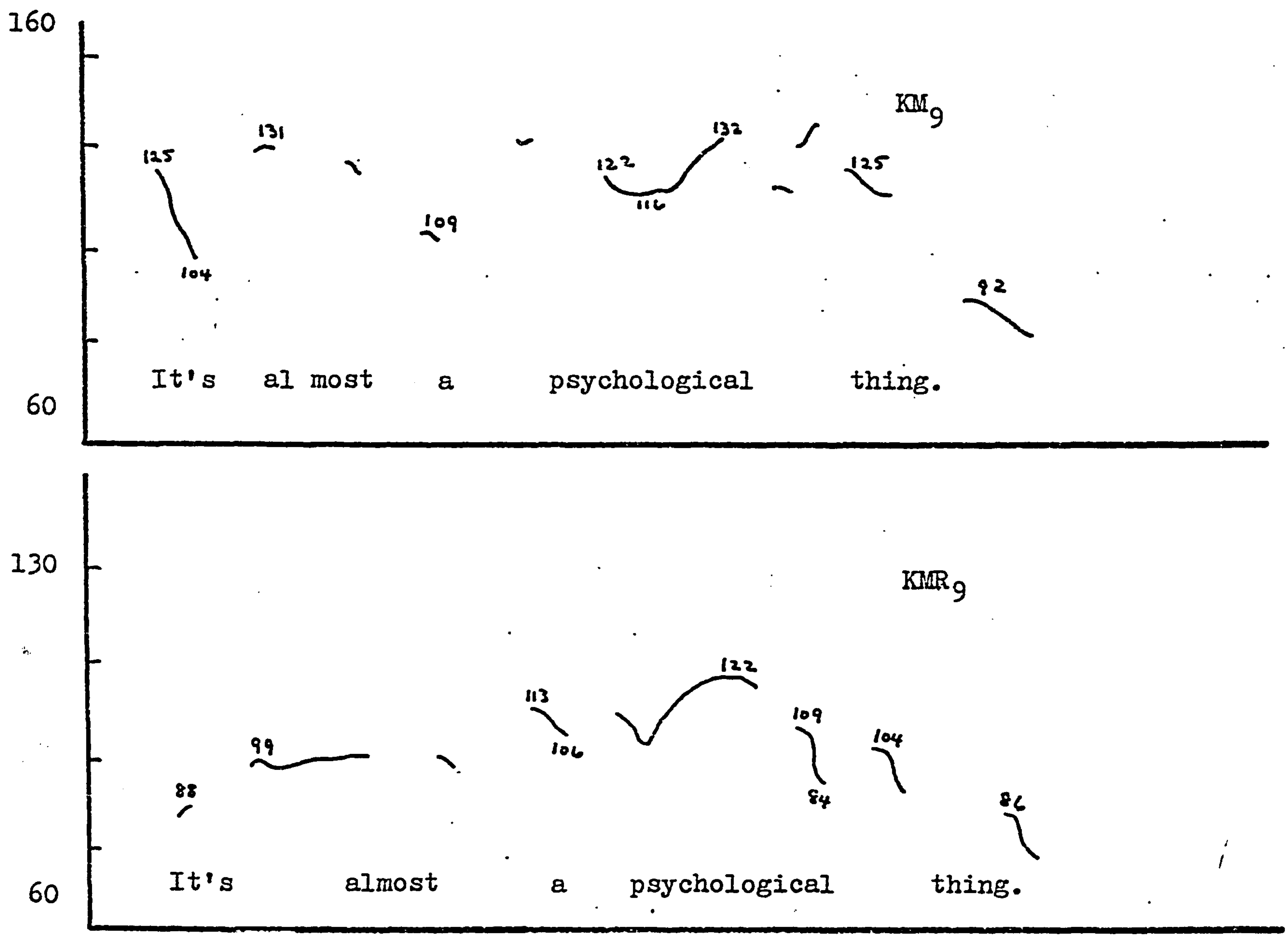

160
60
KM4
this is an a p ti tude test
118
111
116
123
104
84
98
80
123
79


160
60
KMR4
this is an ap ti tude test
118
122
108
120
111
153
108
88
98
92
84
145
99

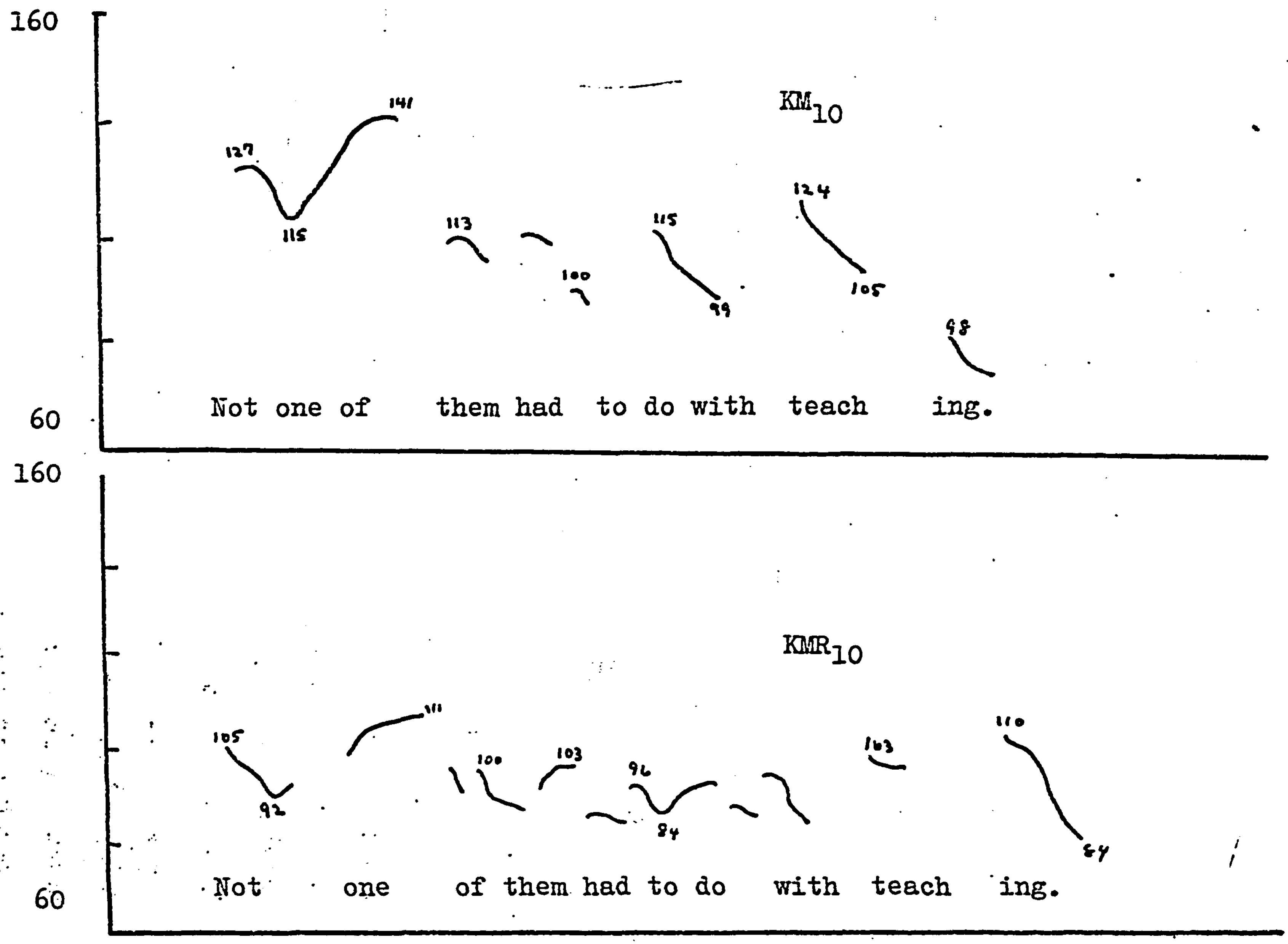

160
KH5
146
120
111
104
105
96
98
94
99
60
It measures inn a t e a bili ty

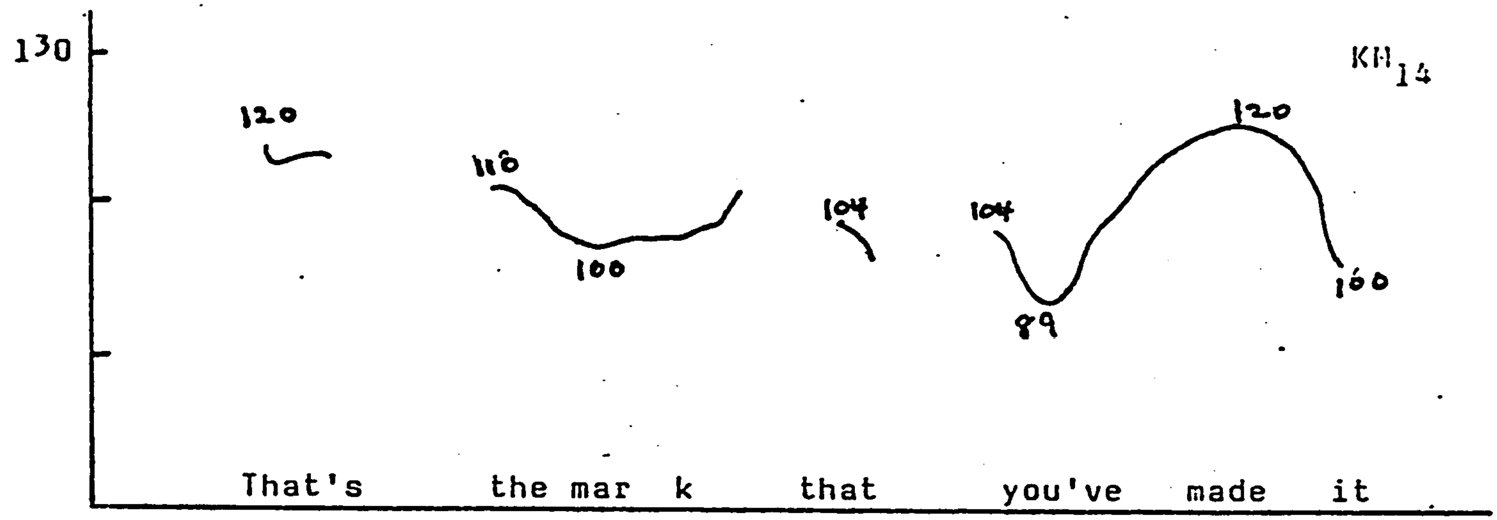

180
160
KHR5
115
110
108
104
96
92
96
90
96
60
It measures innate ability

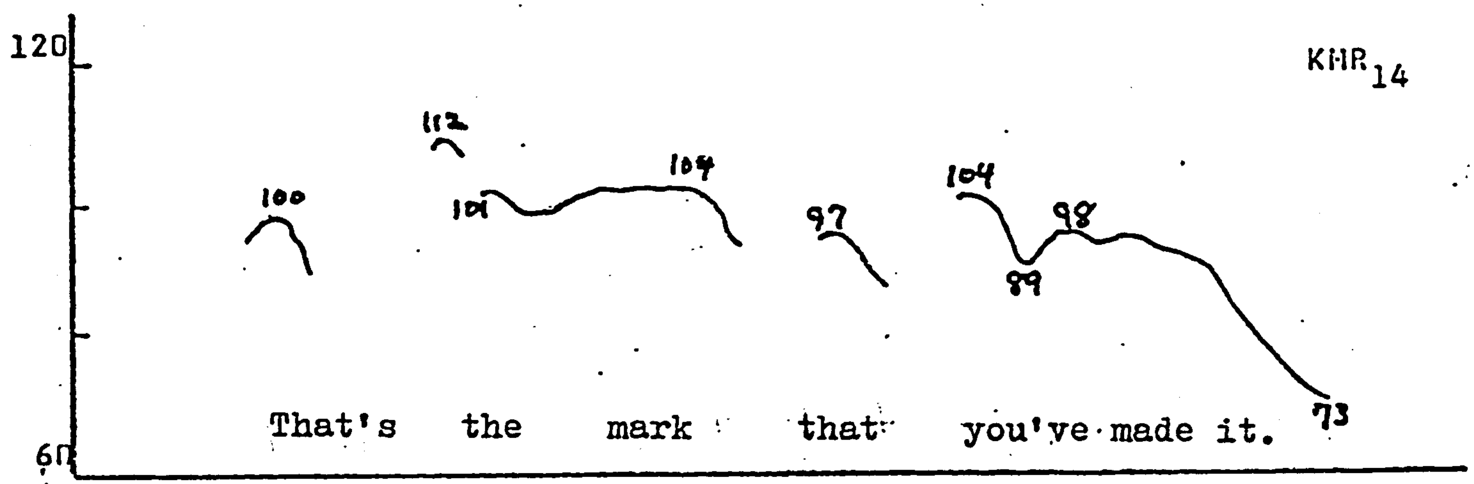

140
KM6
106
90
122
110
98
129
116
101
137
They
tend to lose things literally.
60
120
KMR6
101
112
84
96
112
108
112
92
They tend to lose things literally.
60

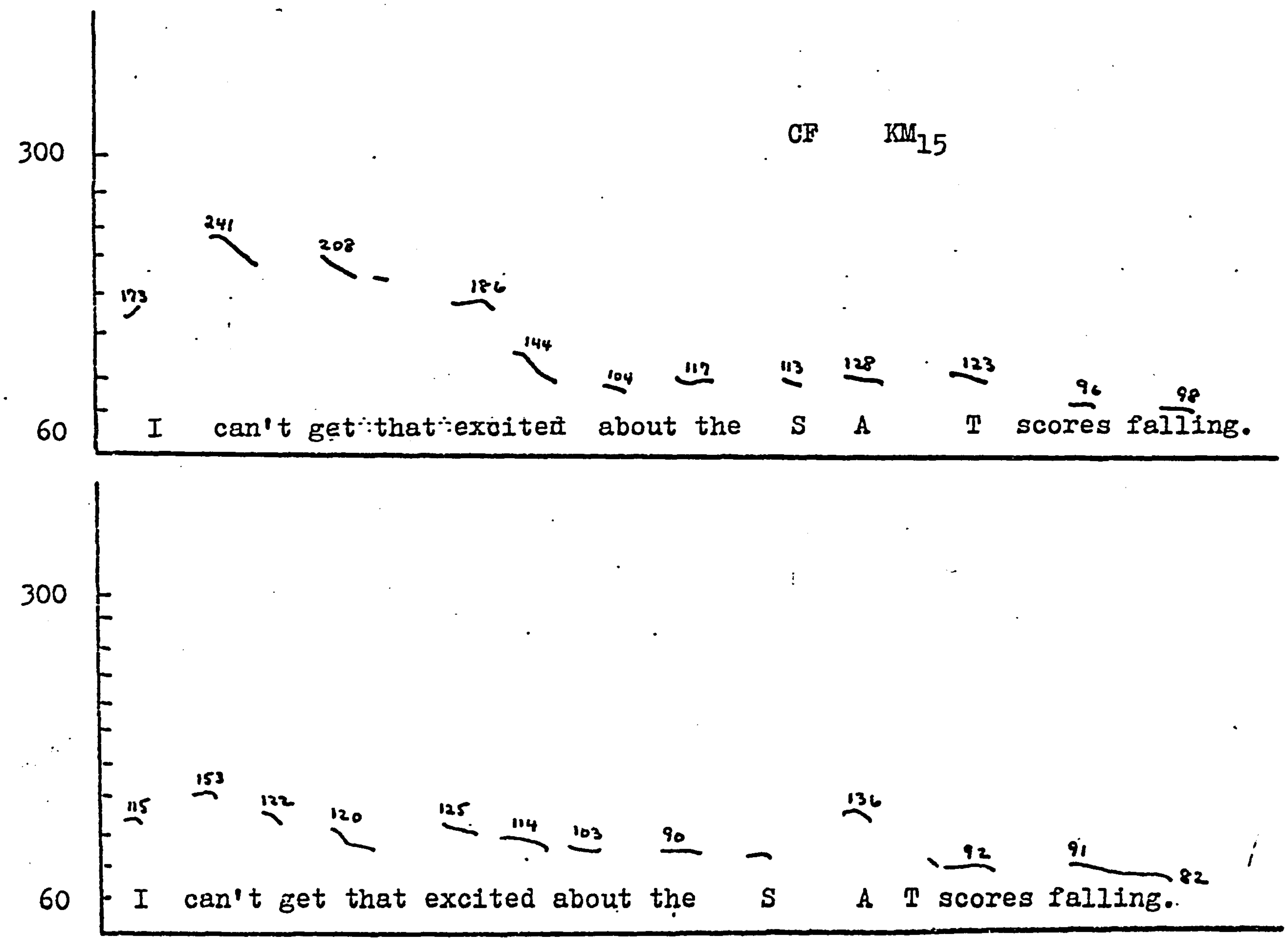


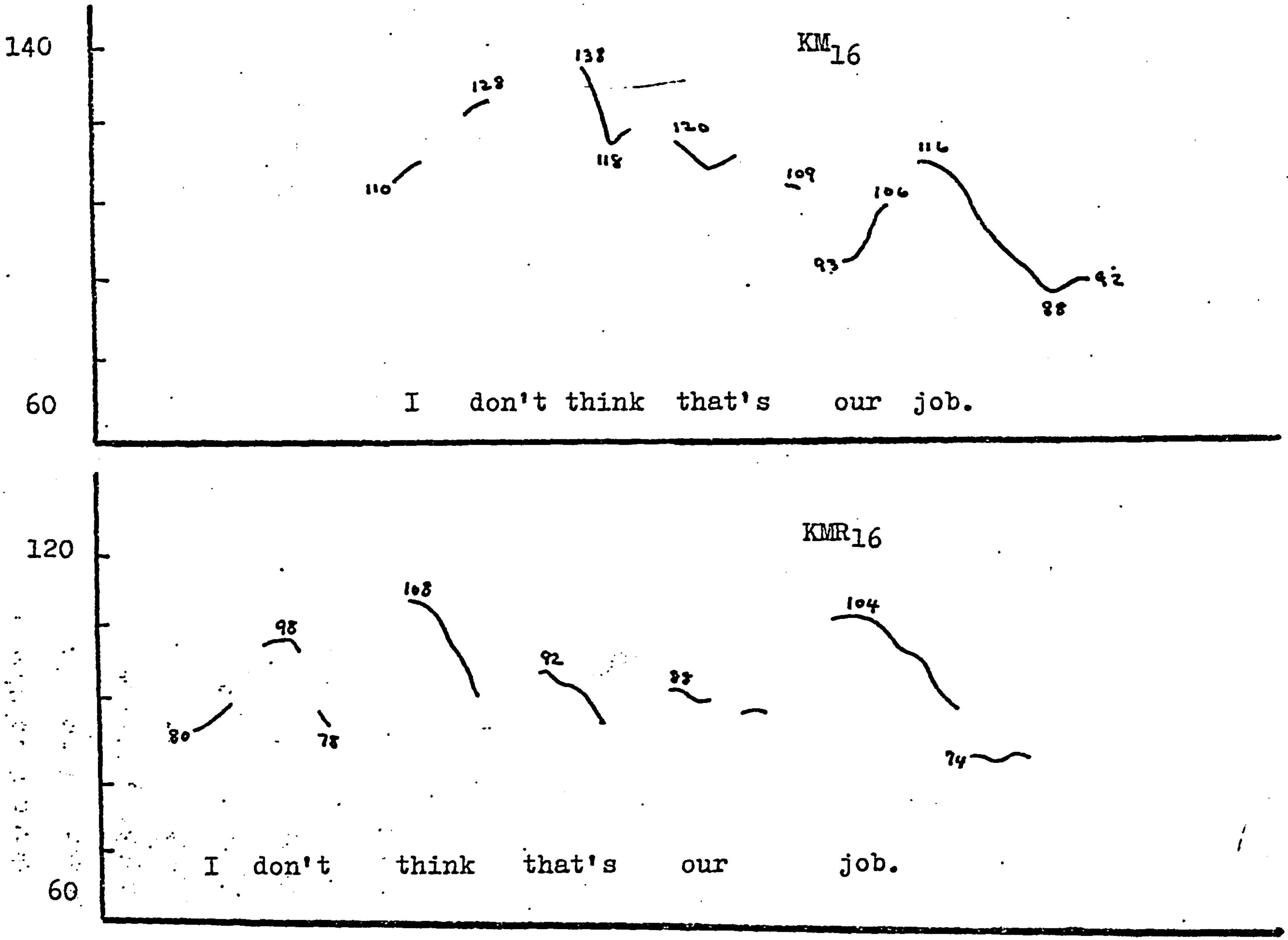
160
60
KM9
125
104
131
109
122
116
132
125
92
It's al most a psychological thing.
130
60
KMR9
88
99
113
106
122
109
84
104
86
It's almost a psychological thing.

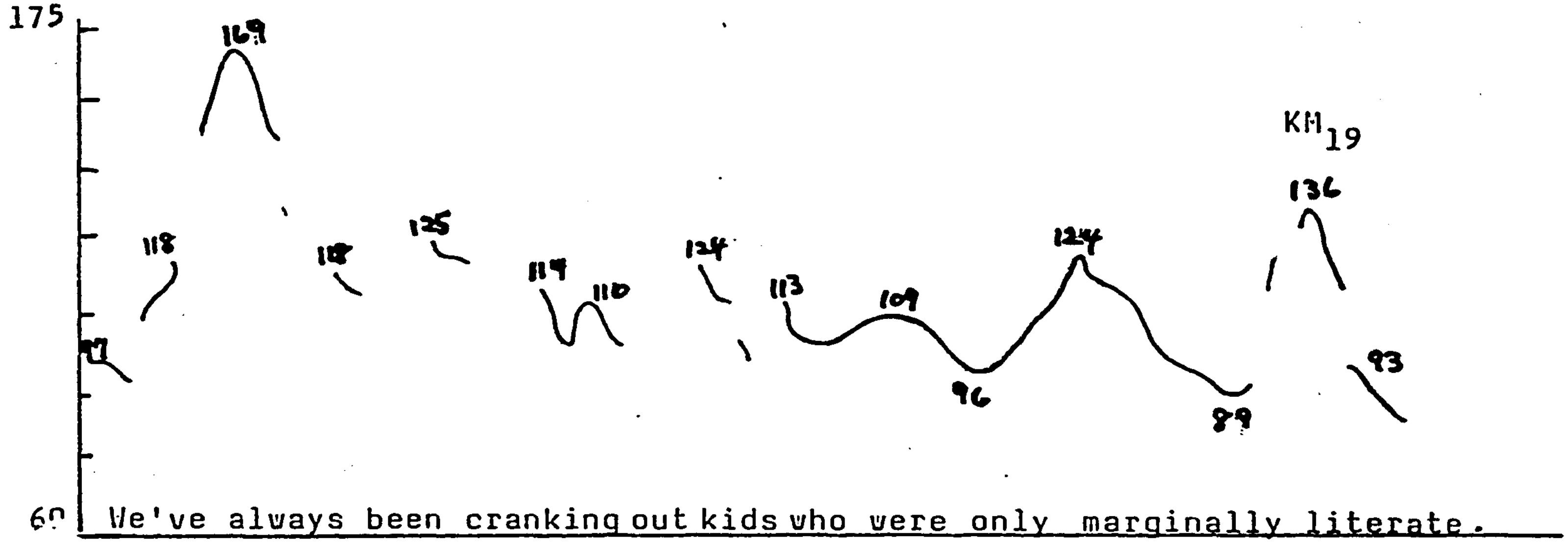

160
KM10
60
Not one of them had to do with teach ing.
160
KMR10
60
Not one one of them had to do with teach ing.

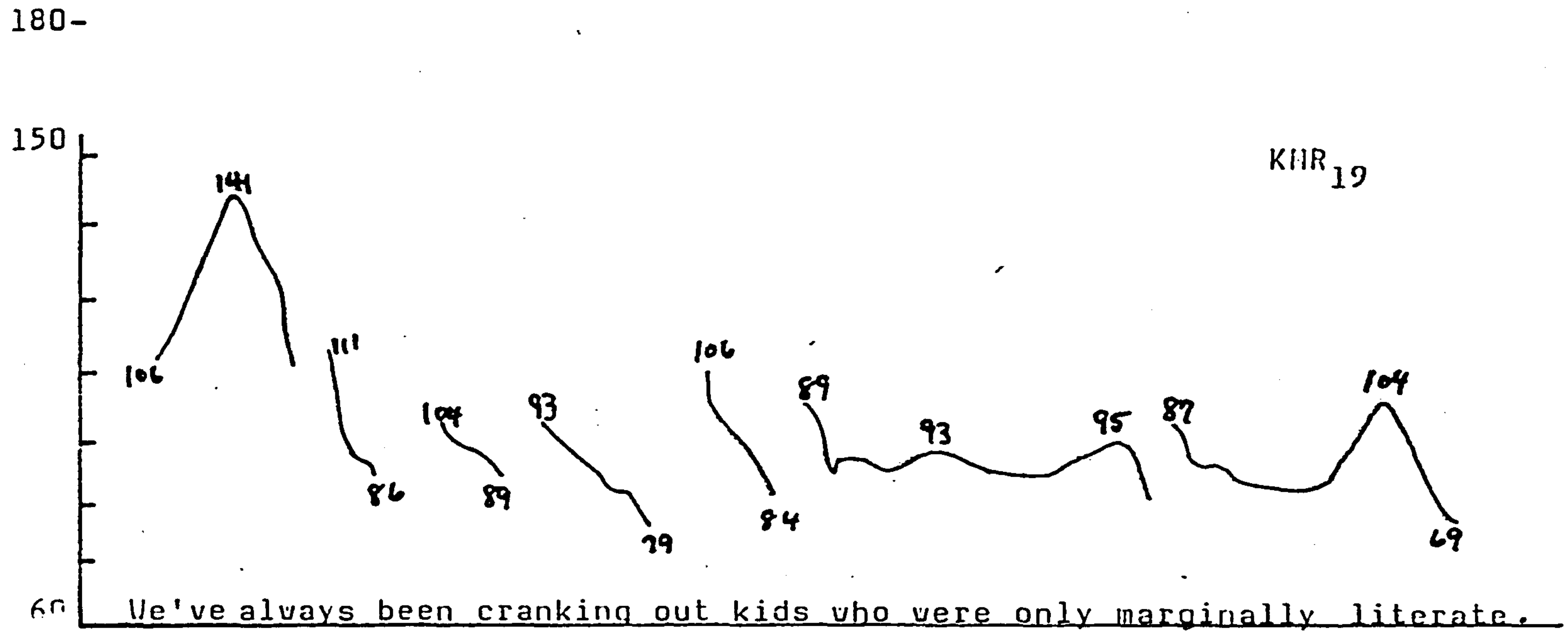

130
KH14
120
118
104
104
120
100
89
100
That's the mar k that you've made it

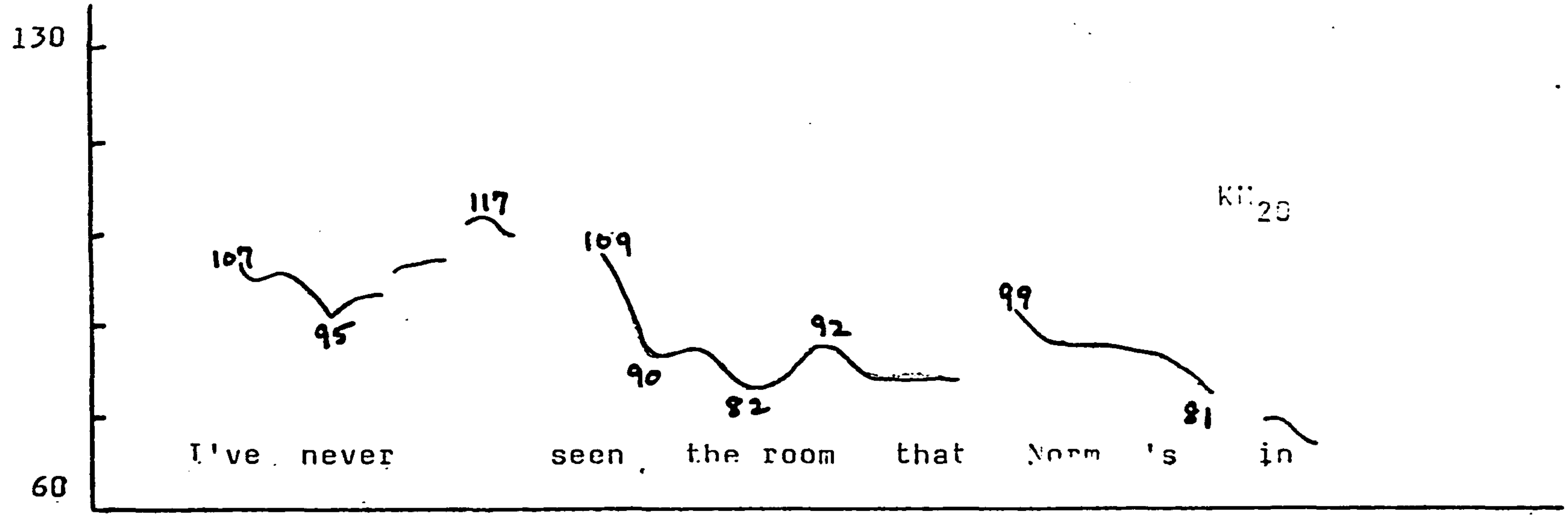

120
KHR14
112
100
101
104
97
104
98
89
73
60
That's the mark that you've made it.

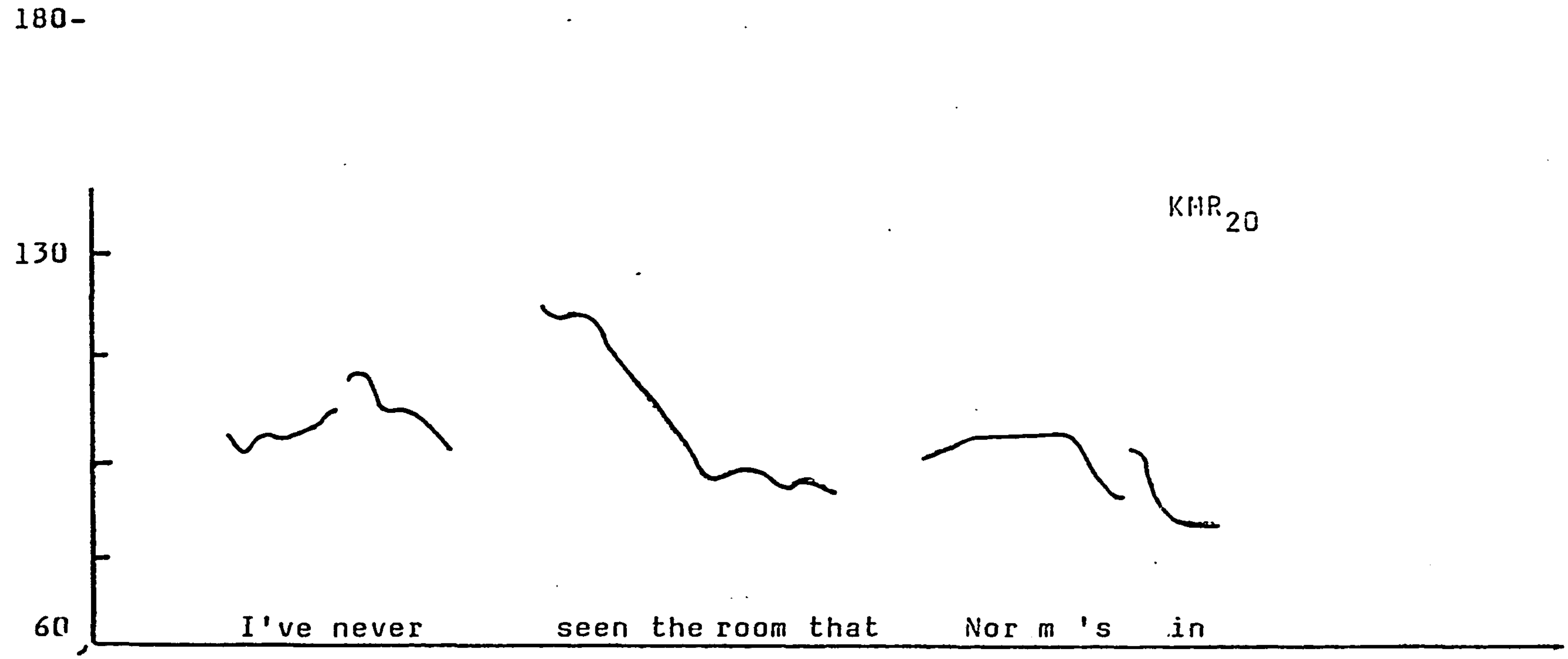

CF KM15
300
60
173 241 208 186 144 104 117 113 128 123 96 98
I can't get that excited about the S A T scores falling.
300
60
115 153 122 120 125 114 103 90 136 92 91 82
I can't get that excited about the S A T scores falling.

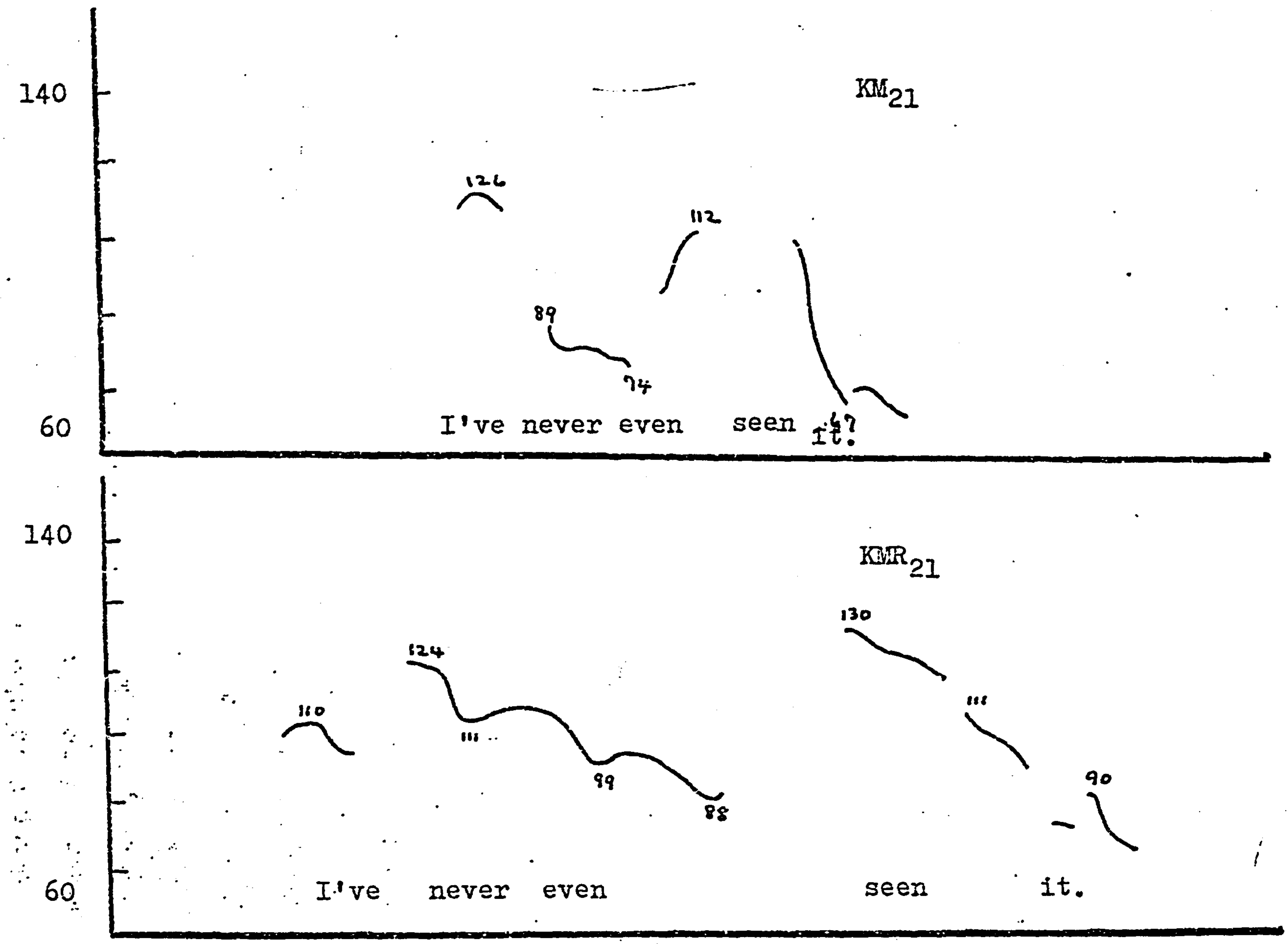

140
KM16
138
128
120
118
116
110
109
106
93
92
88
60
I don't think that's our job.
120
KMR16
108
104
98
92
88
80
78
74
60
I don't think that's our job.

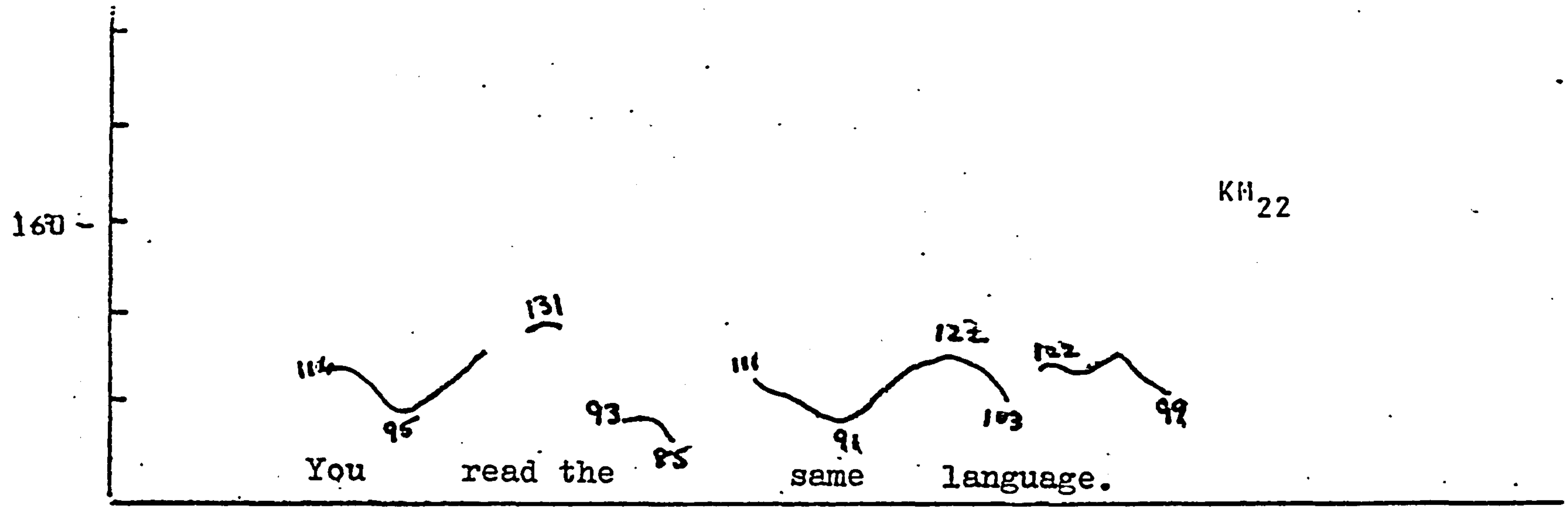

175
60
KM_{19}
169
118
118
125
119
110
124
113
109
96
124
89
136
93
We've always been cranking out kids who were only marginally literate.

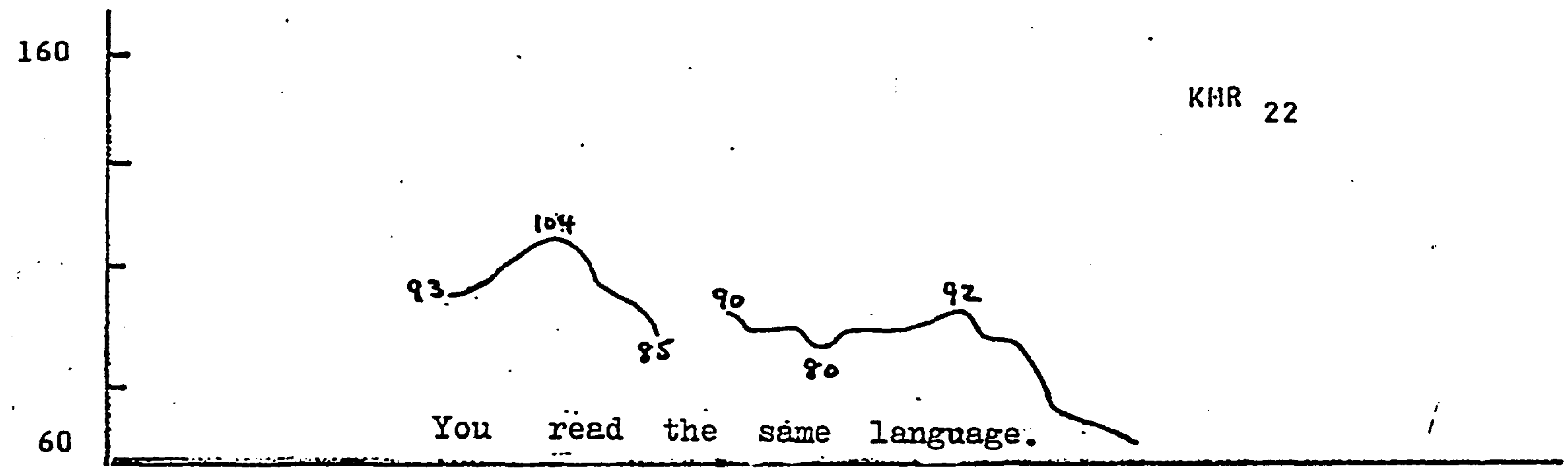

180-
150
60
KMR_{19}
106
141
111
86
104
89
93
79
106
84
89
93
95
87
104
69
We've always been cranking out kids who were only marginally literate.

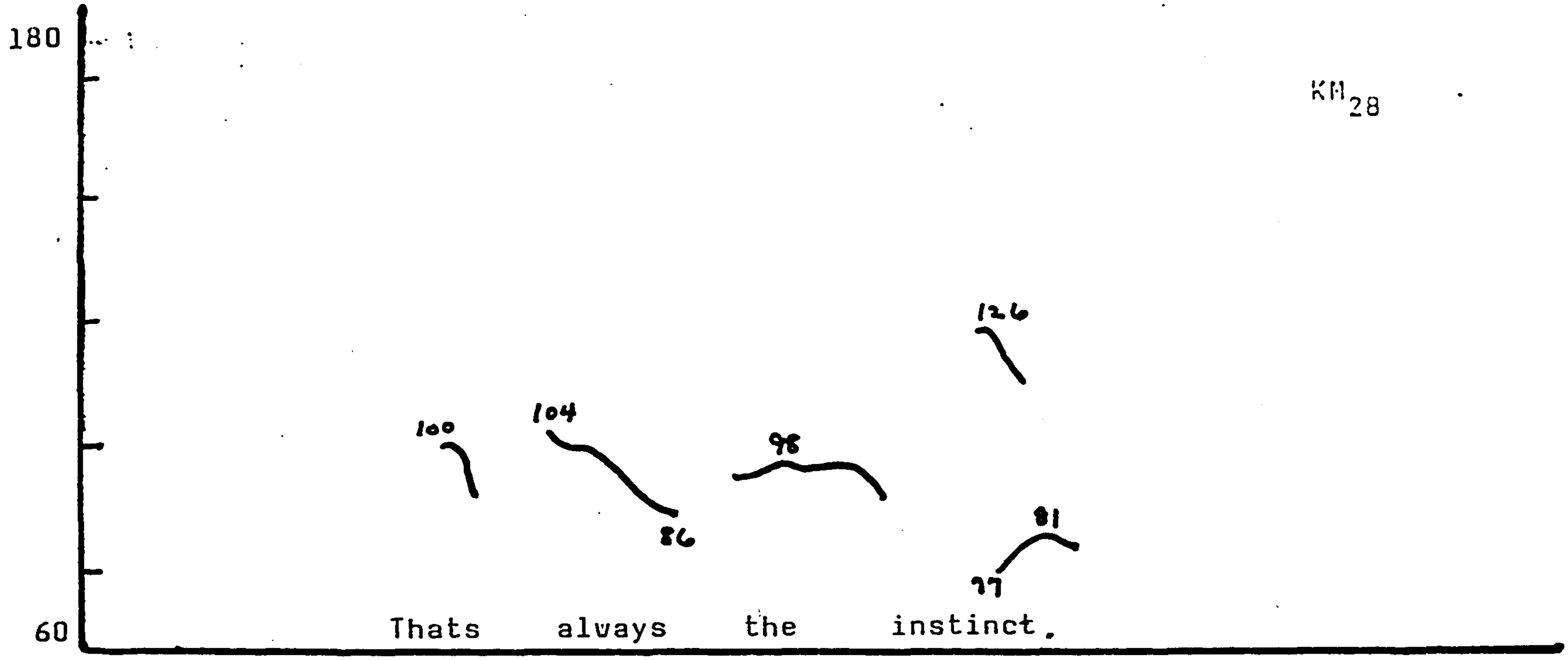

130
60
KM20
107
95
117
109
90
82
92
99
81
I've never seen the room that Norm 's in

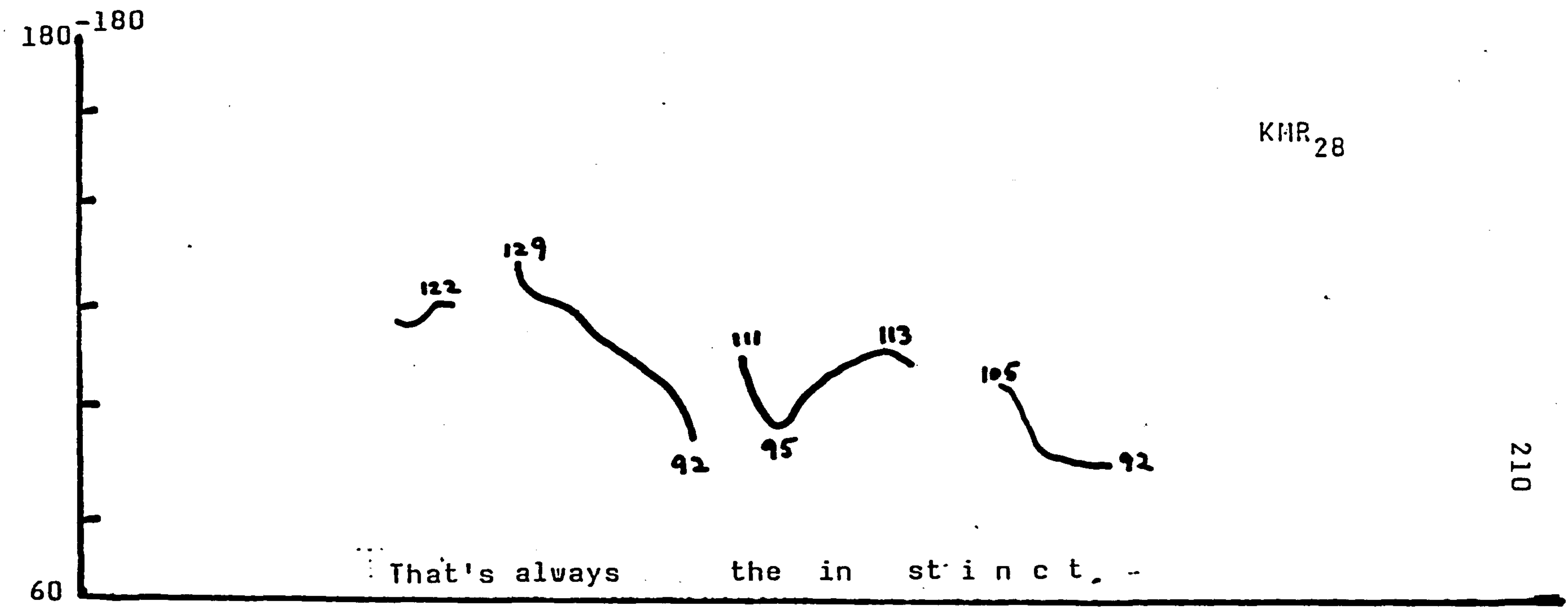

180-
130
60
KMR20
I've never seen the room that Nor m 's in

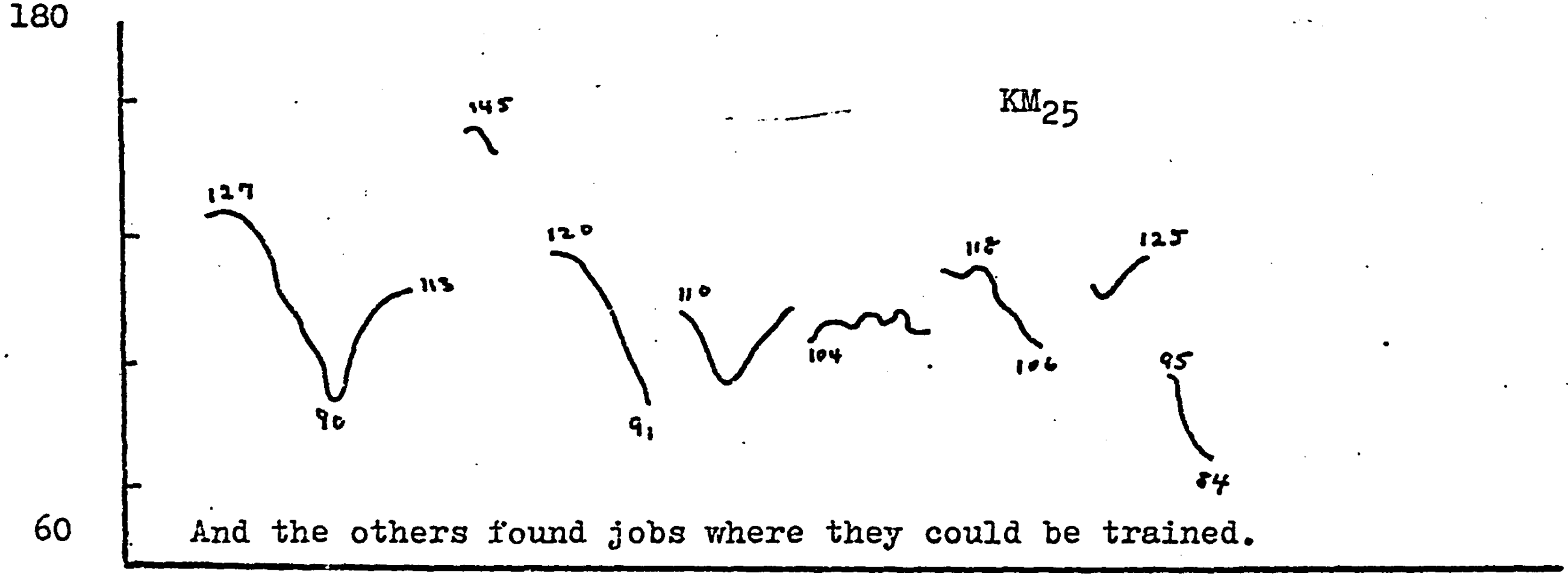

140
60
KM21
126
112
89
74
67
I've never even seen it.
140
60
KMR21
130
124
110
111
111
99
88
90
I've never even seen it.

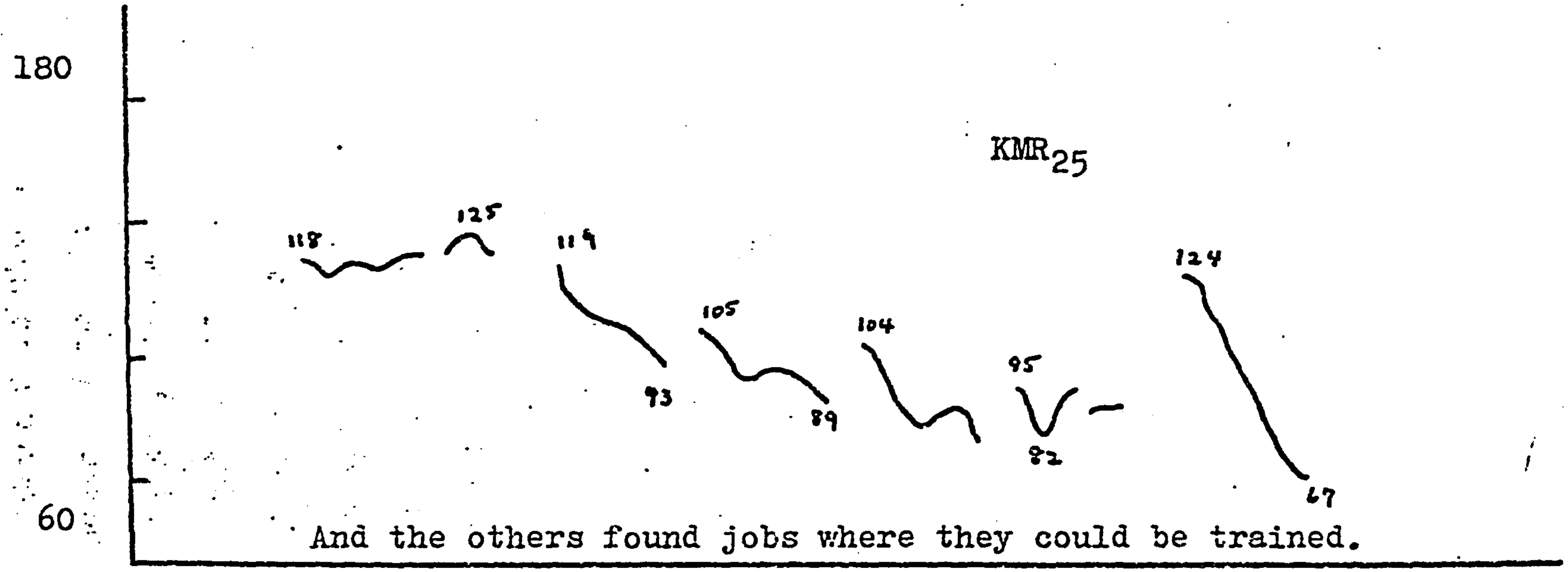

160
KH 22
111
95
131
93
85
111
91
122
103
122
99
You read the same language.

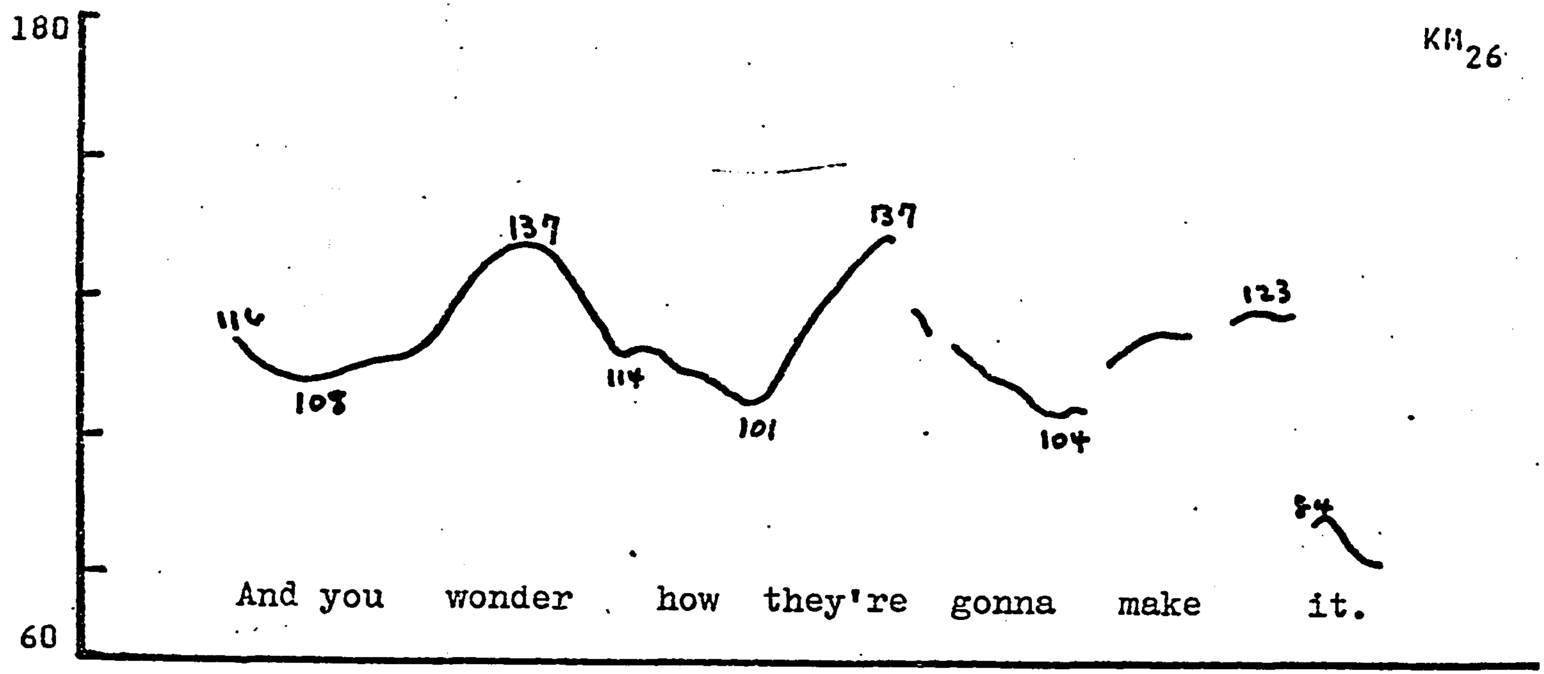

160
KHR 22
93
104
85
90
80
92
60
You read the same language.

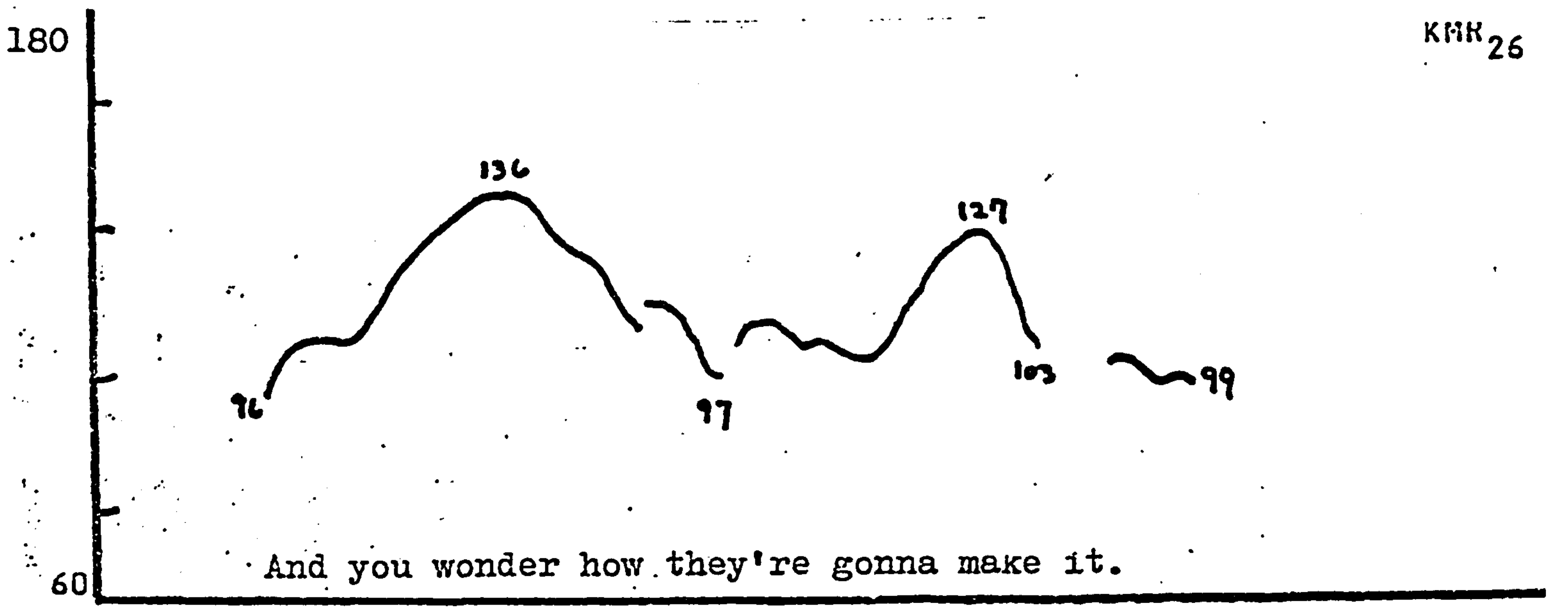

180
60
KM28
100
104
86
98
126
77
81
Thats always the instinct.

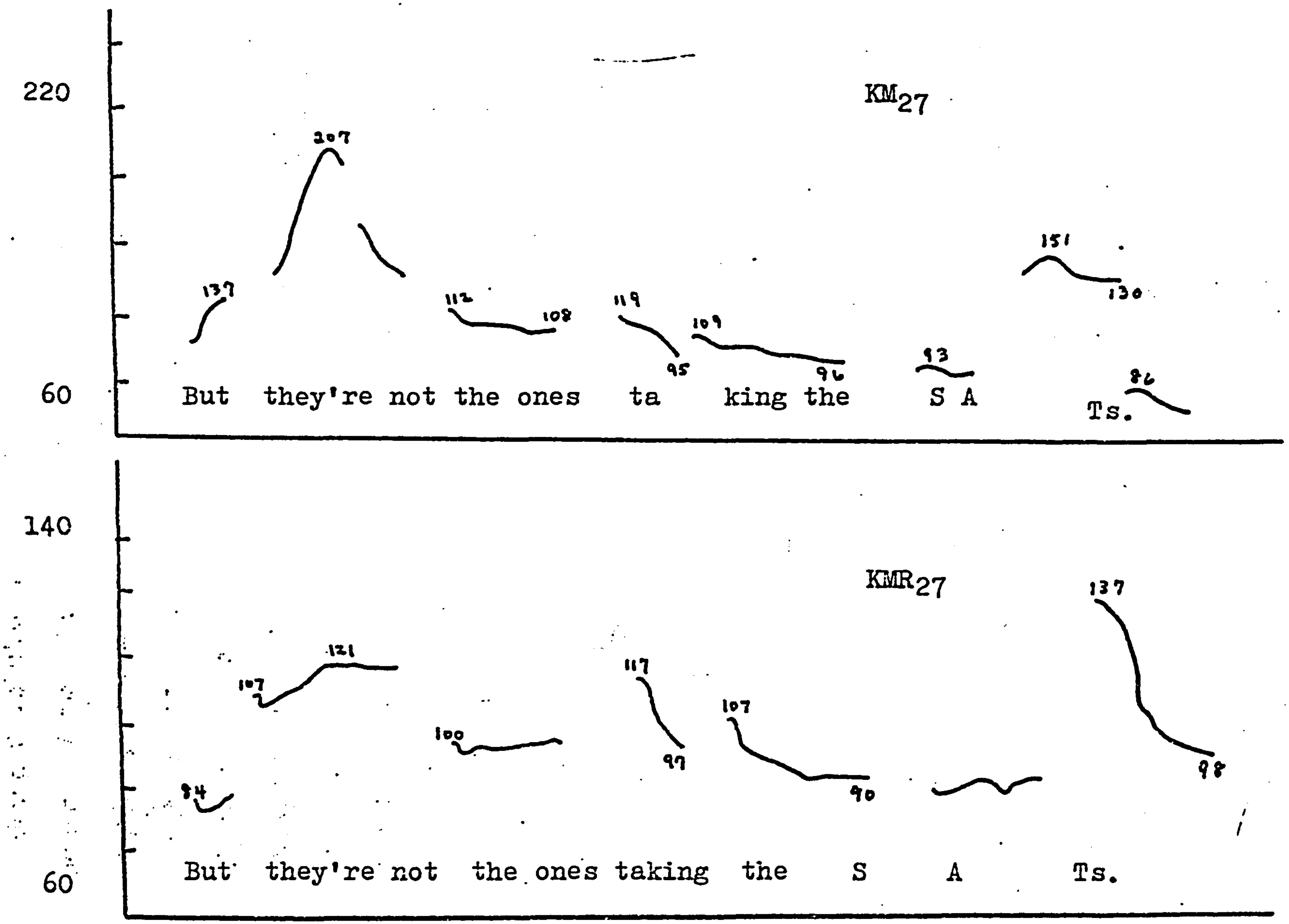

180
-180
60
KMR28
122
129
92
111
95
113
105
92
That's always the in st i n c t.


180
60
KM25
145
127
113
90
120
91
110
104
118
106
125
95
84
And the others found jobs where they could be trained.

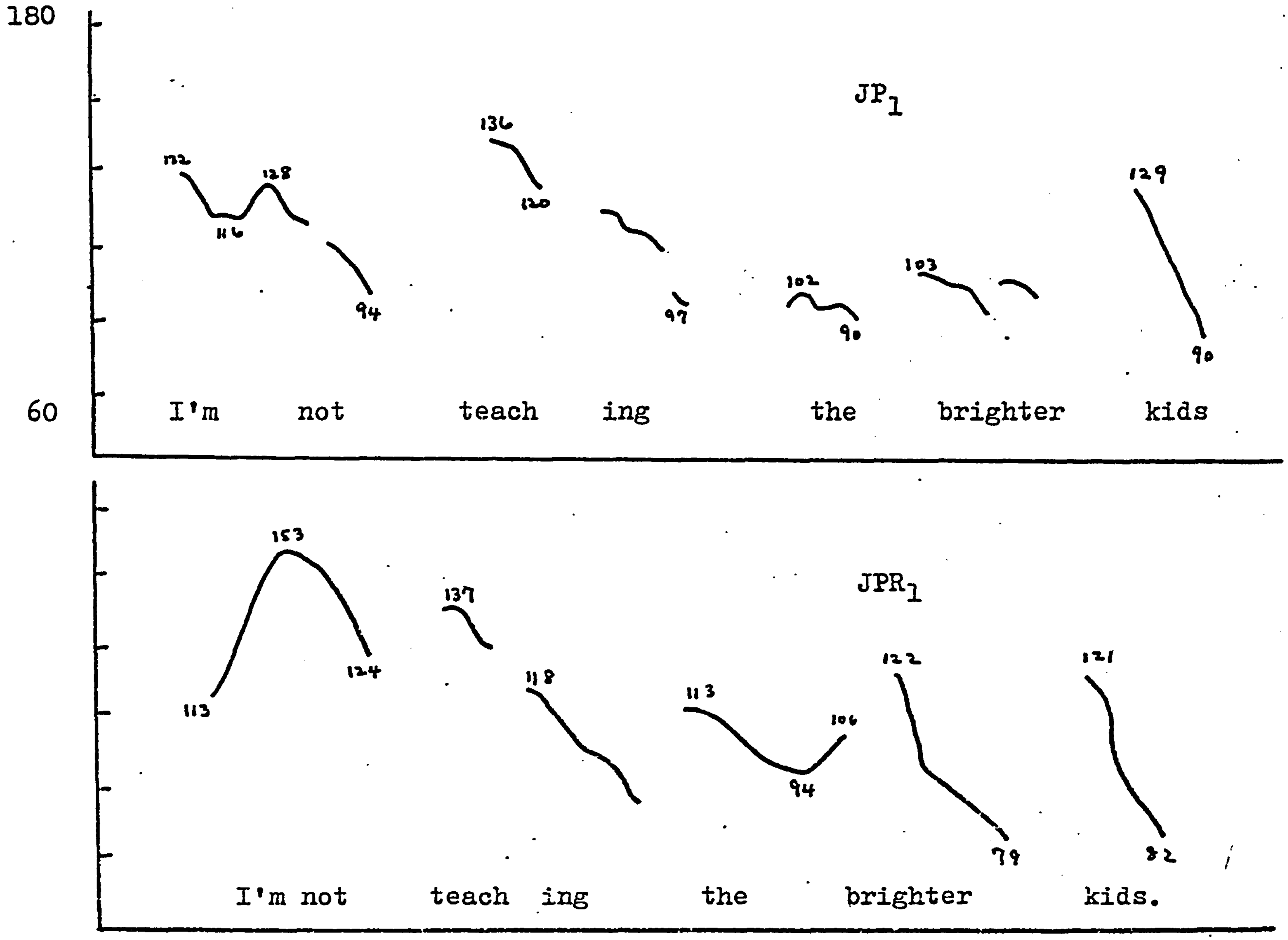

180
60
KMR25
118
125
119
93
105
89
104
95
82
124
67
And the others found jobs where they could be trained.

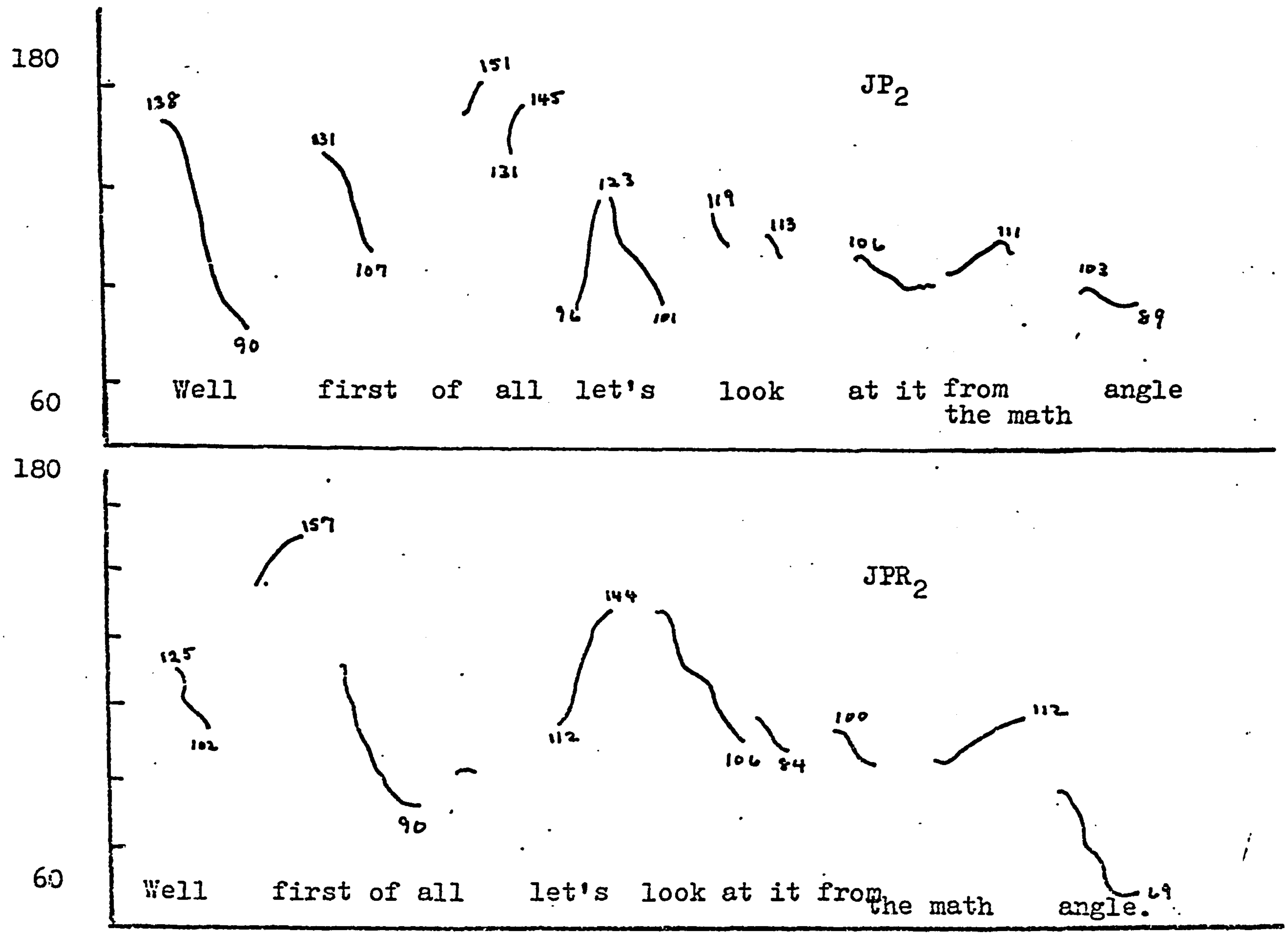

180
60
KM26
116
108
137
114
101
137
104
123
84
And you wonder how they're gonna make it.

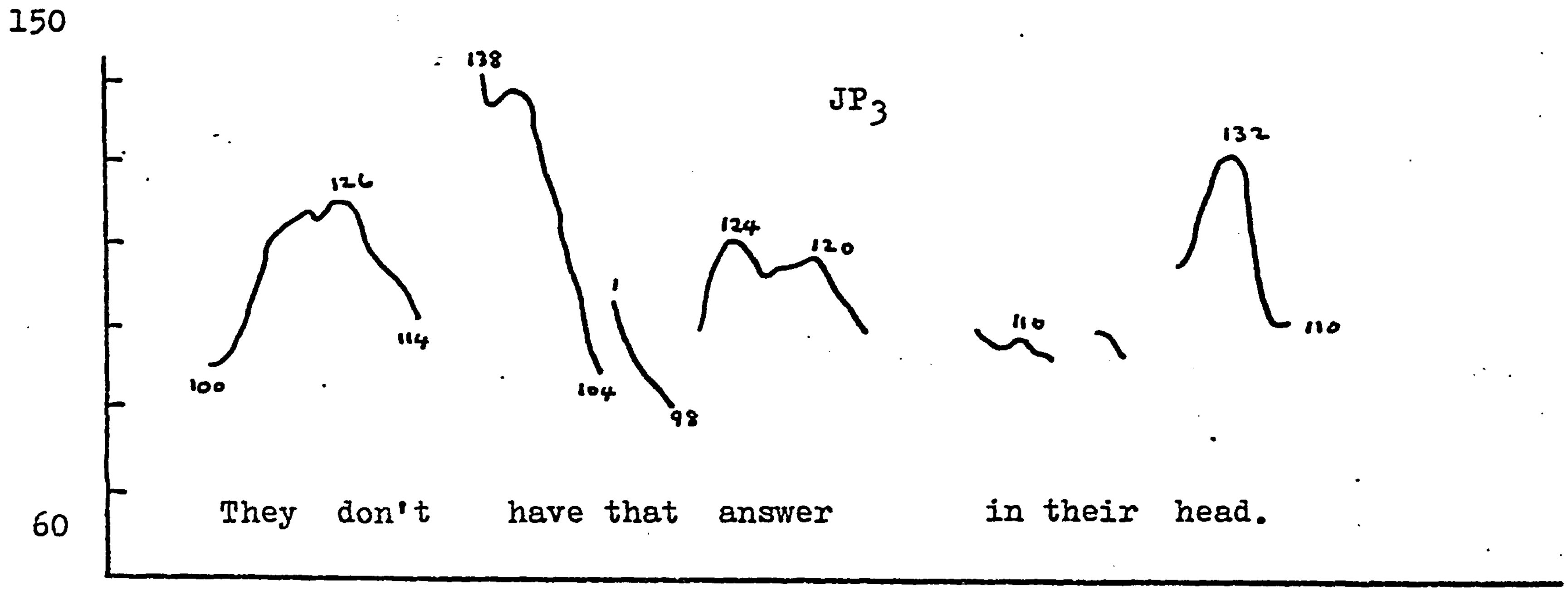

180
60
KMK26
96
136
97
127
103
99
And you wonder how they're gonna make it.

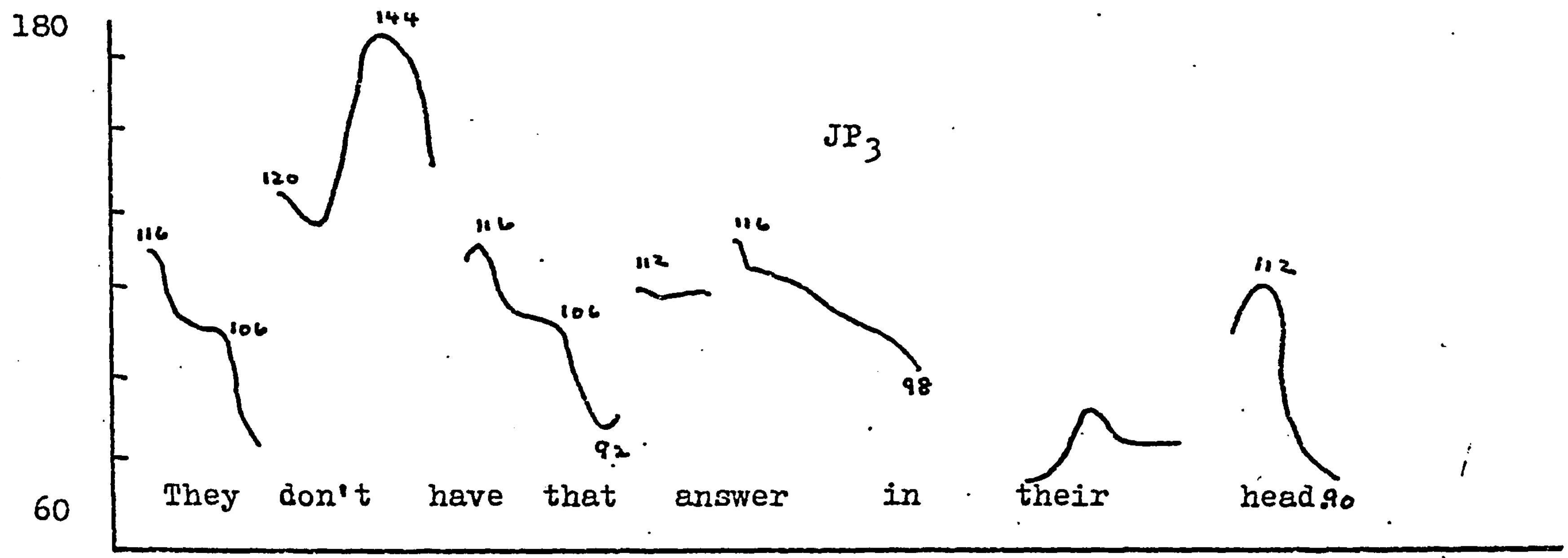


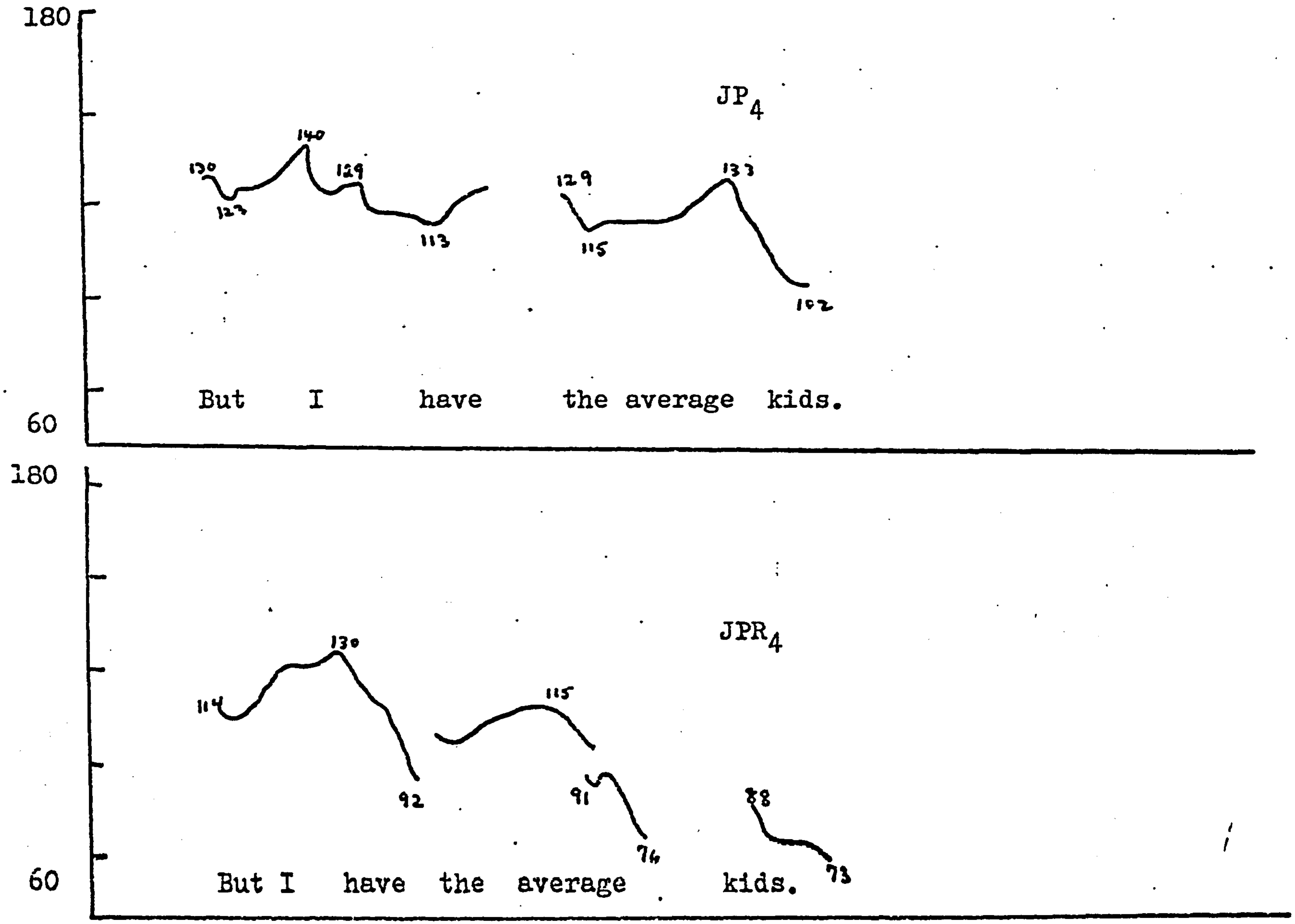
220
60
KM27
207
137
112
108
119
95
109
96
93
151
130
86
But they're not the ones ta king the S A Ts.
140
60
KMR27
121
107
84
100
117
97
107
90
137
98
But they're not the ones taking the S A Ts.

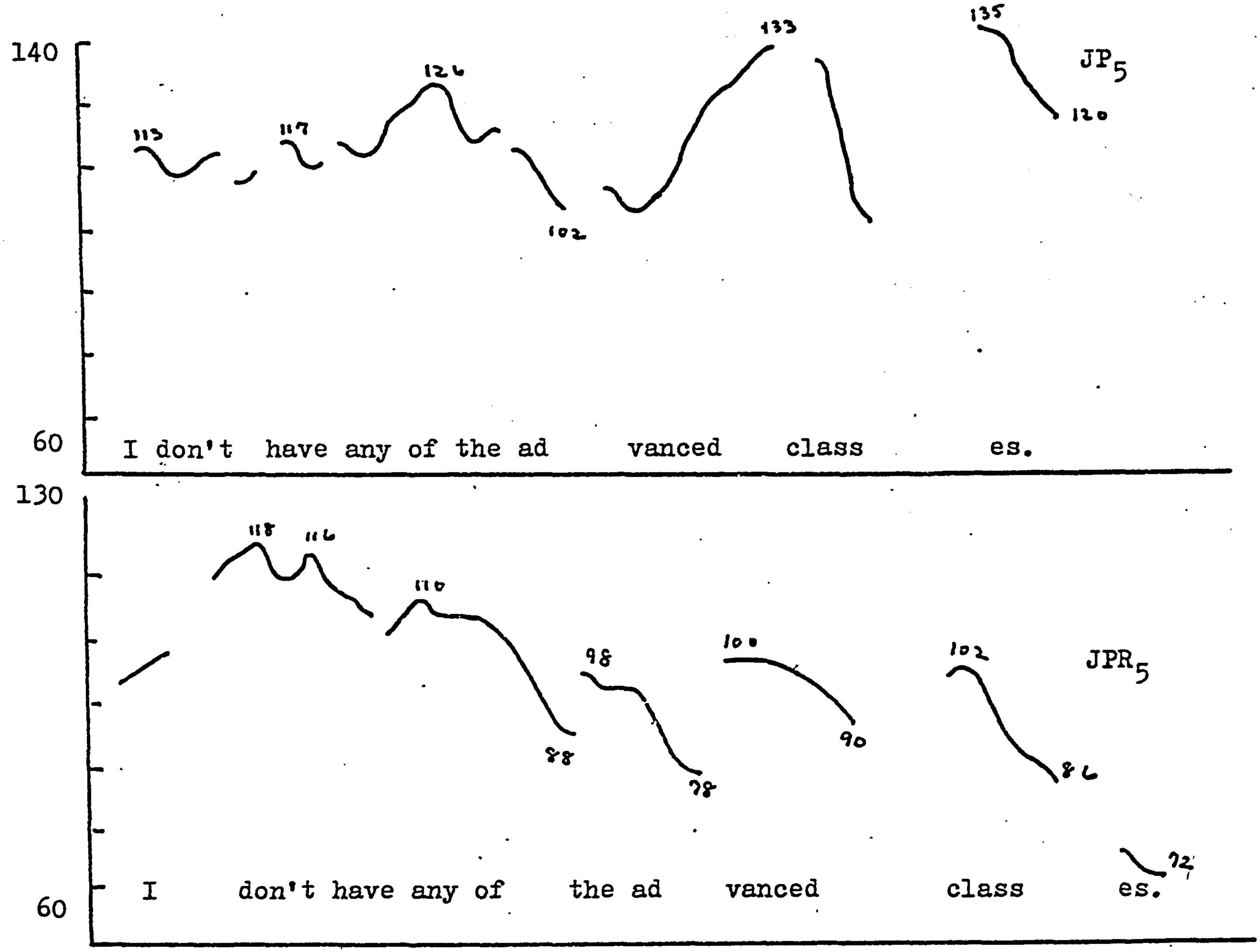

180
60
JP1
132
116
128
94
136
120
97
102
90
103
129
90
I'm not teach ing the brighter kids
JPR1
153
113
124
137
118
113
94
106
122
79
121
82
I'm not teach ing the brighter kids.

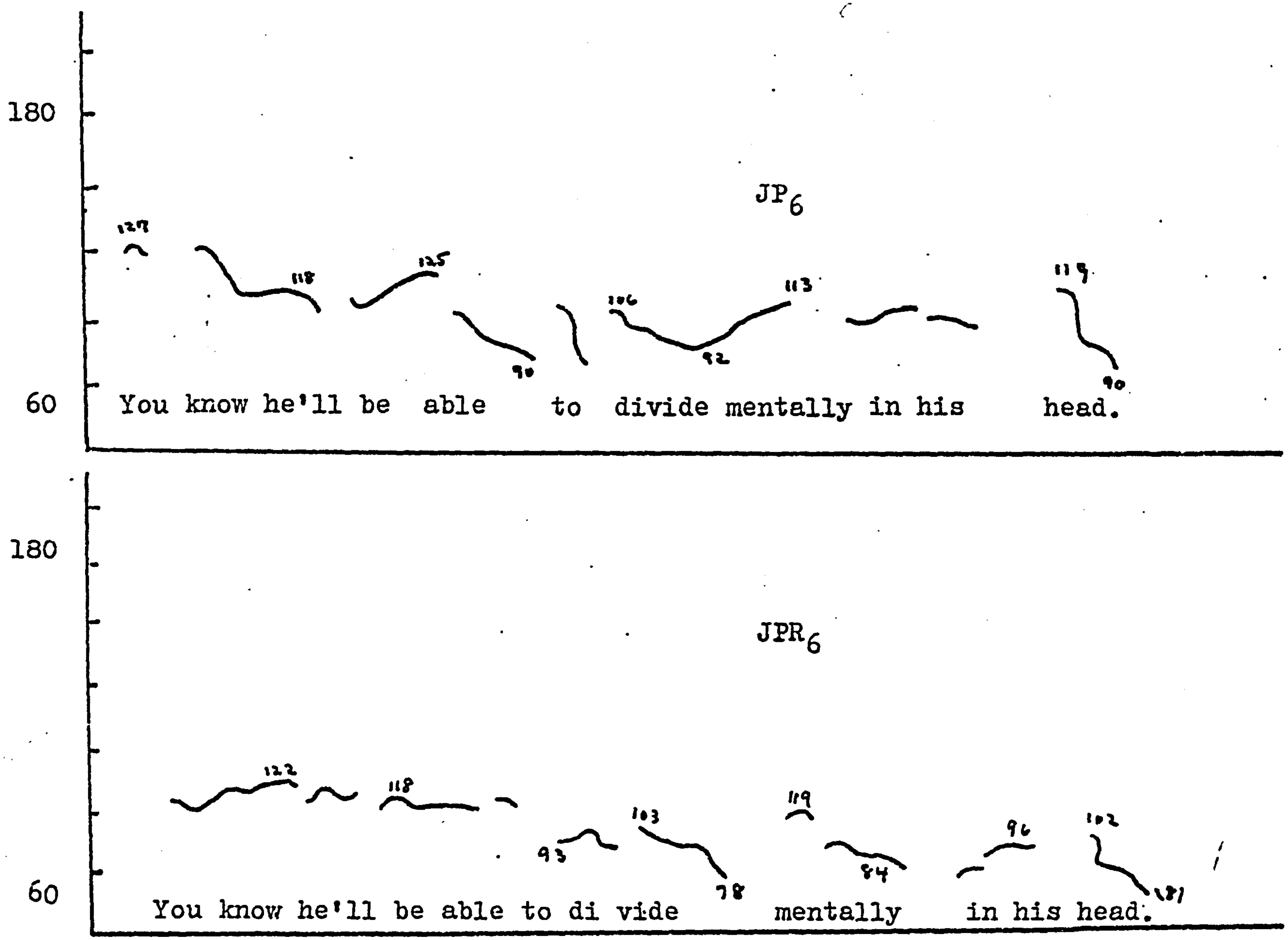

180
60
JP2
138
90
131
107
151
145
131
123
96
101
119
113
106
111
103
89
Well first of all let's look at it from the math angle
180
60
JPR2
125
102
157
90
144
112
106
84
100
112
69
Well first of all let's look at it from the math angle.

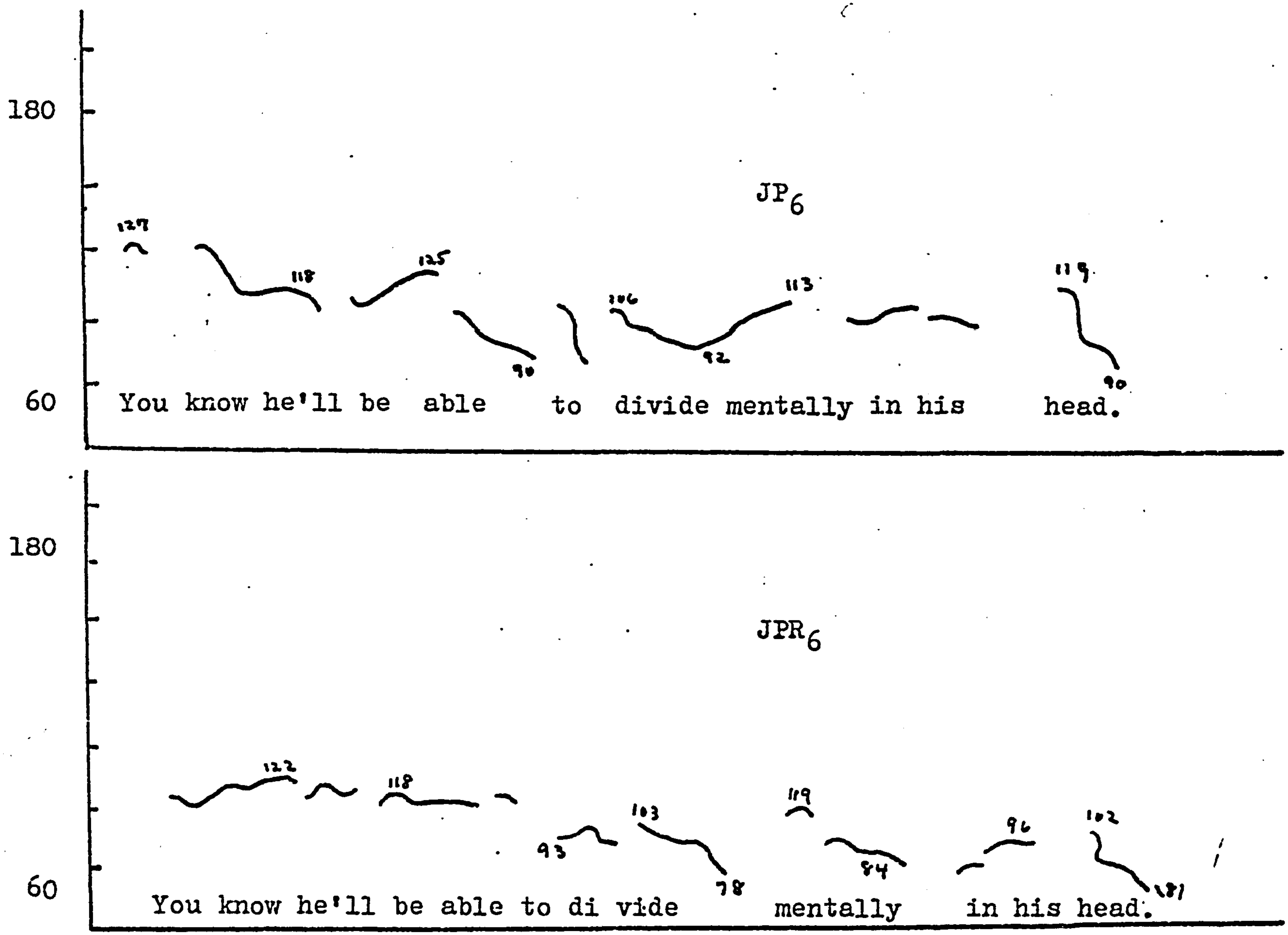

150
60
JP3
138
126
124
120
132
100
114
104
98
110
110
They don't have that answer in their head.

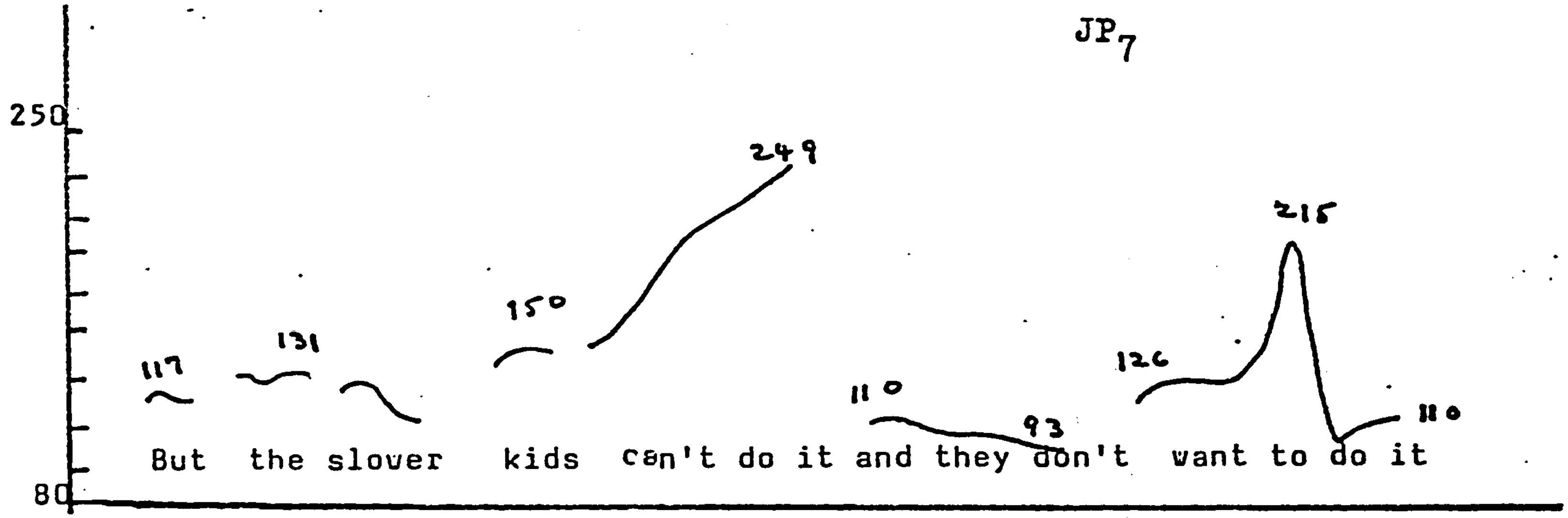

180
60
JP3
144
120
116
106
116
106
92
112
116
98
112
They don't have that answer in their head 90

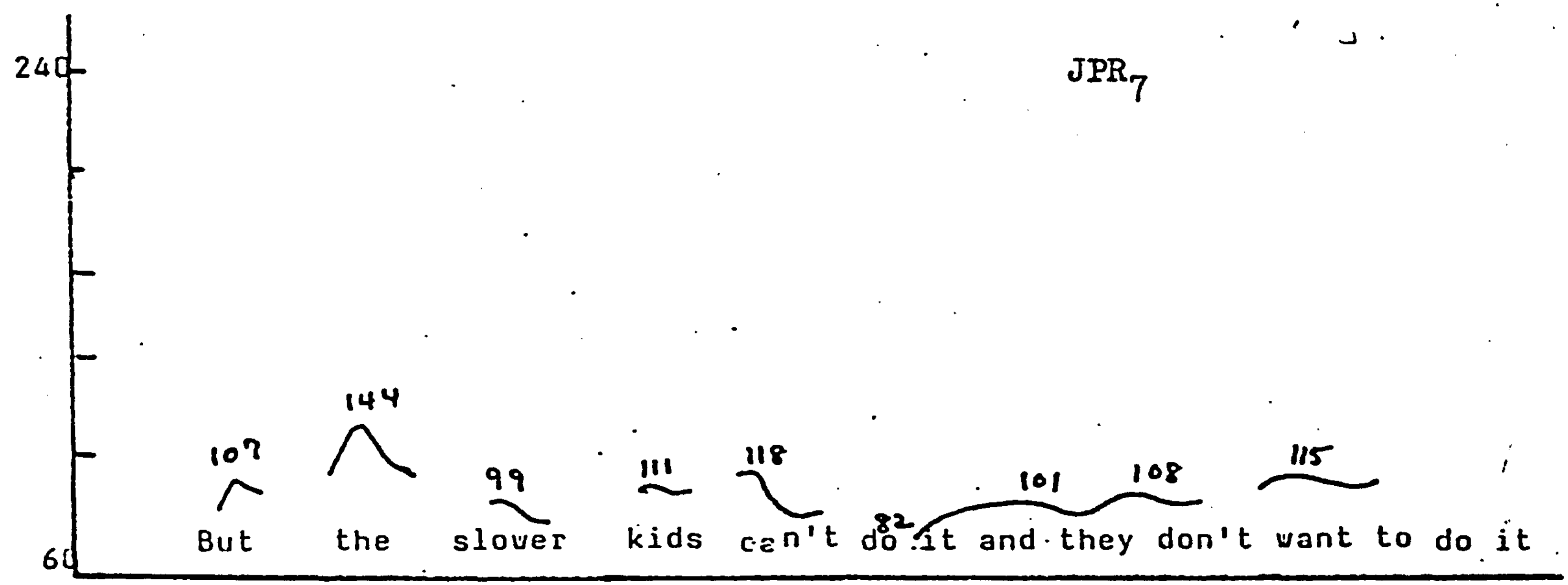

180
60
JP$_4$
130
123
140
129
113
129
115
133
102
But I have the average kids.
180
60
JPR$_4$
114
130
92
115
91
74
88
73
But I have the average kids.

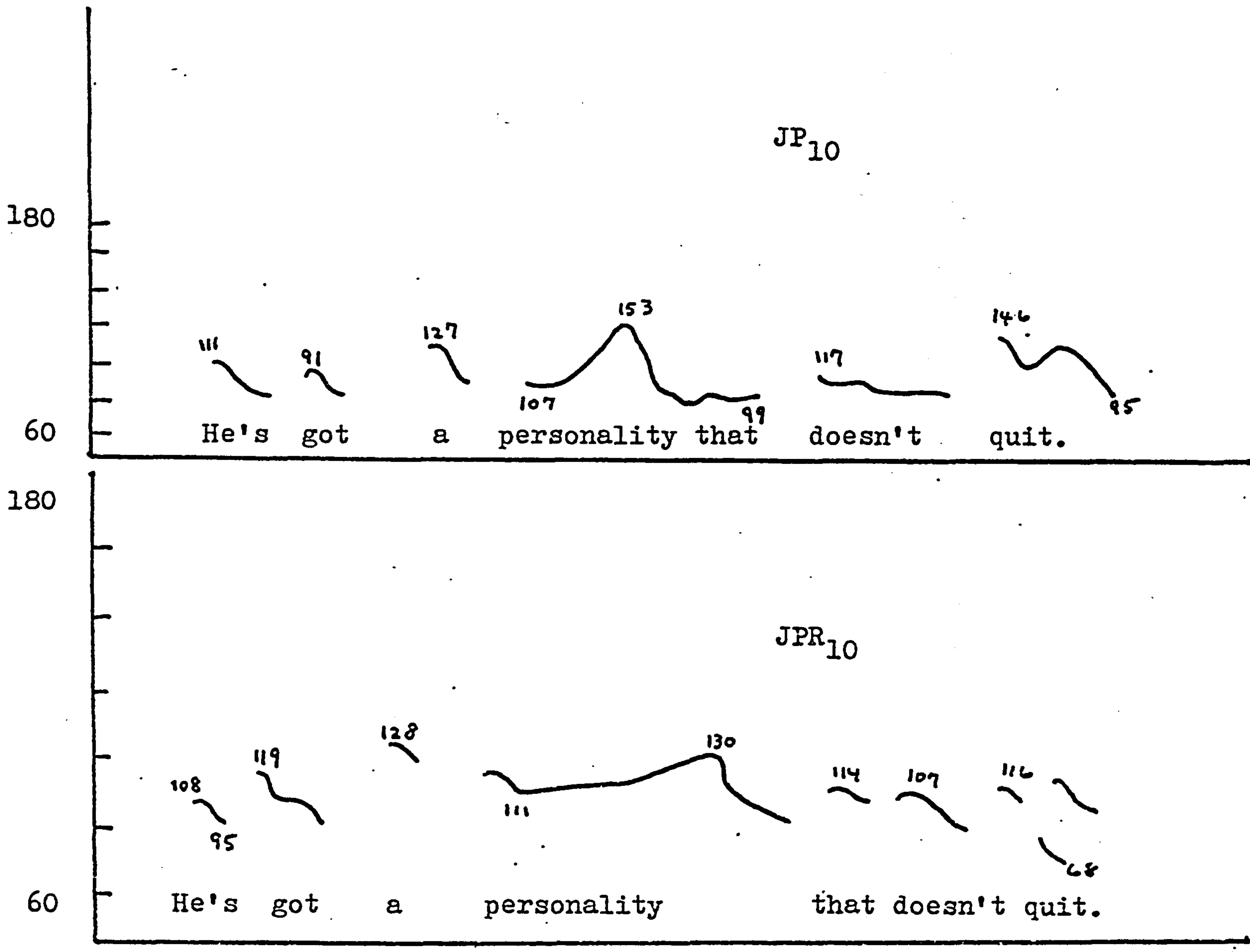

140
60
113
117
126
102
133
135
120
JP5
I don't have any of the ad vanced class es.
130
60
118
116
110
88
98
78
100
90
102
86
72
JPR5
I don't have any of the ad vanced class es.


180
60
JP6
You know he'll be able to divide mentally in his head.
180
60
JPR6
You know he'll be able to di vide mentally in his head.

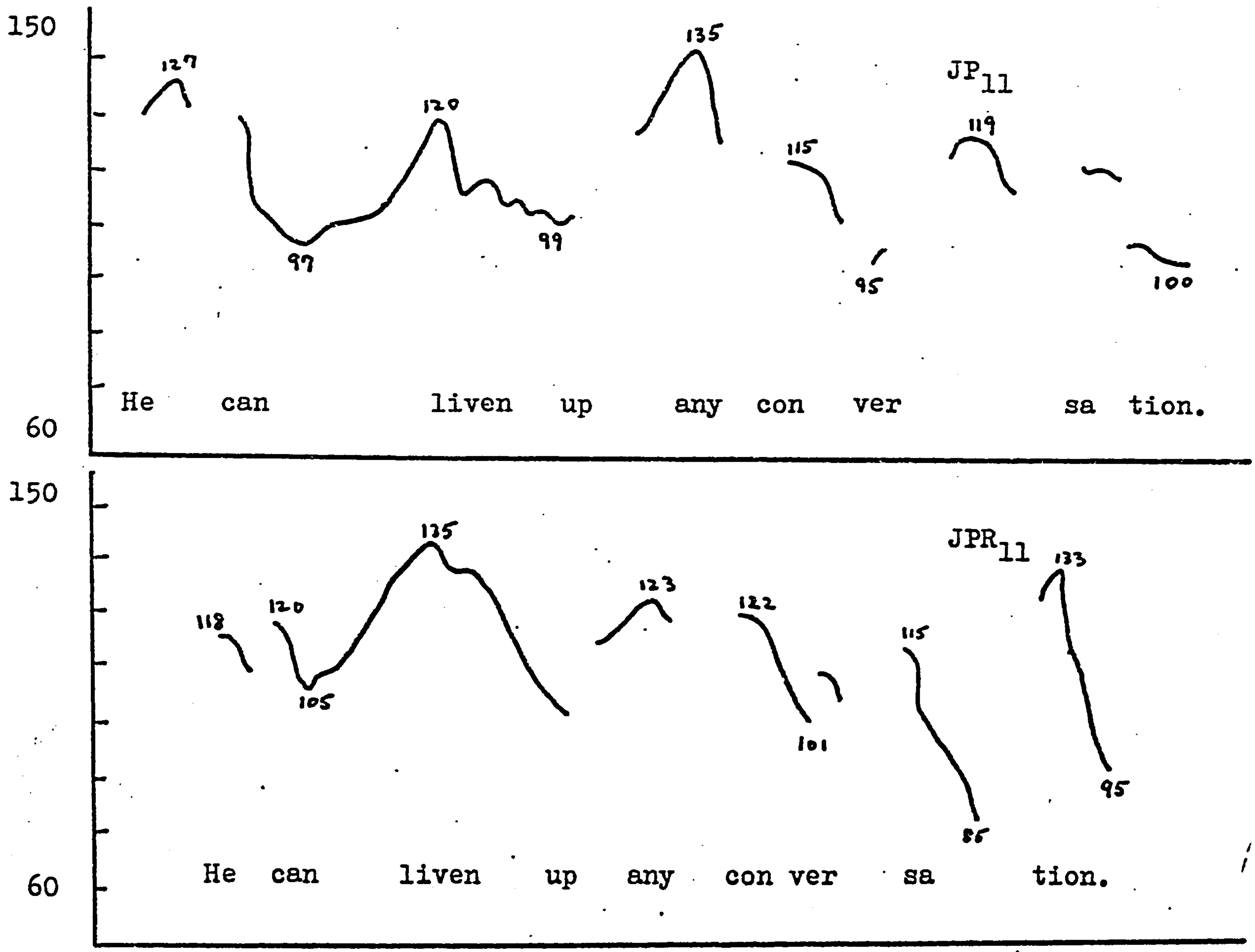

180
60
JP6
127
118
125
90
106
92
113
119
90
You know he'll be able to divide mentally in his head.
180
60
JPR6
122
118
93
103
78
119
84
96
102
81
You know he'll be able to di vide mentally in his head.

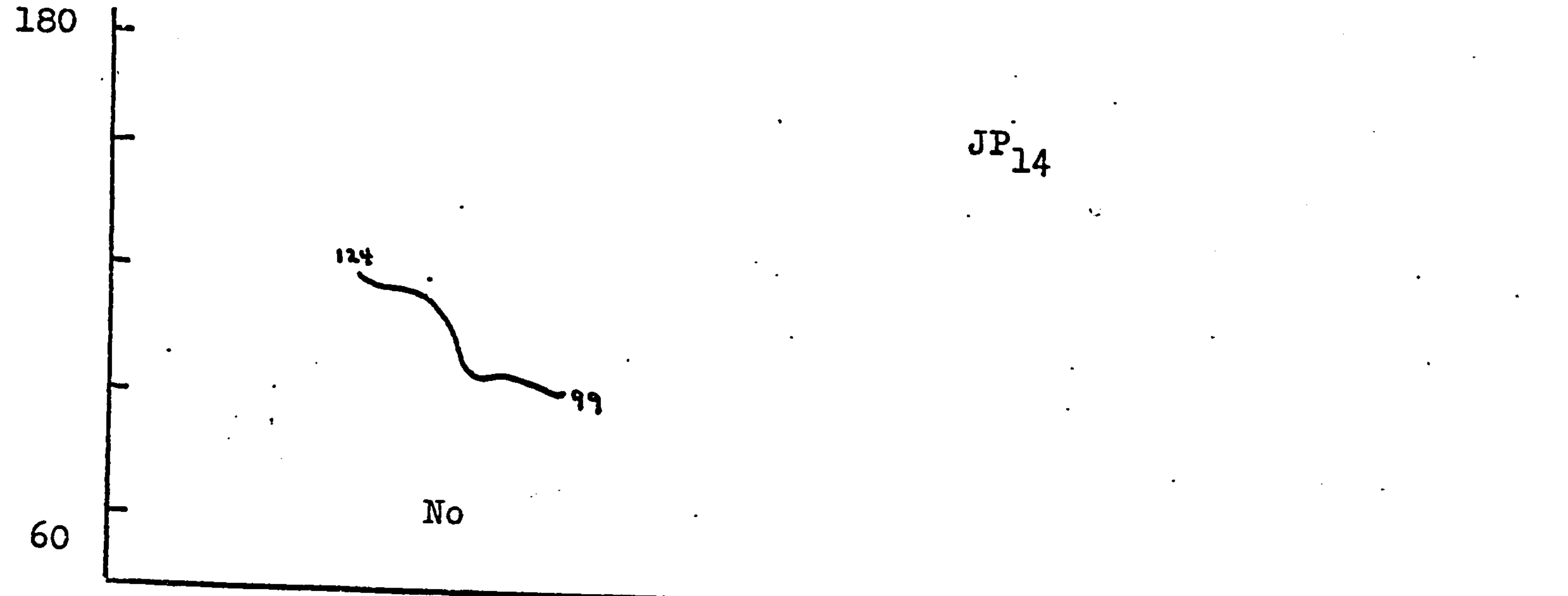

JP7
250
80
117
131
150
249
110
93
126
215
110
But the slower kids can't do it and they don't want to do it

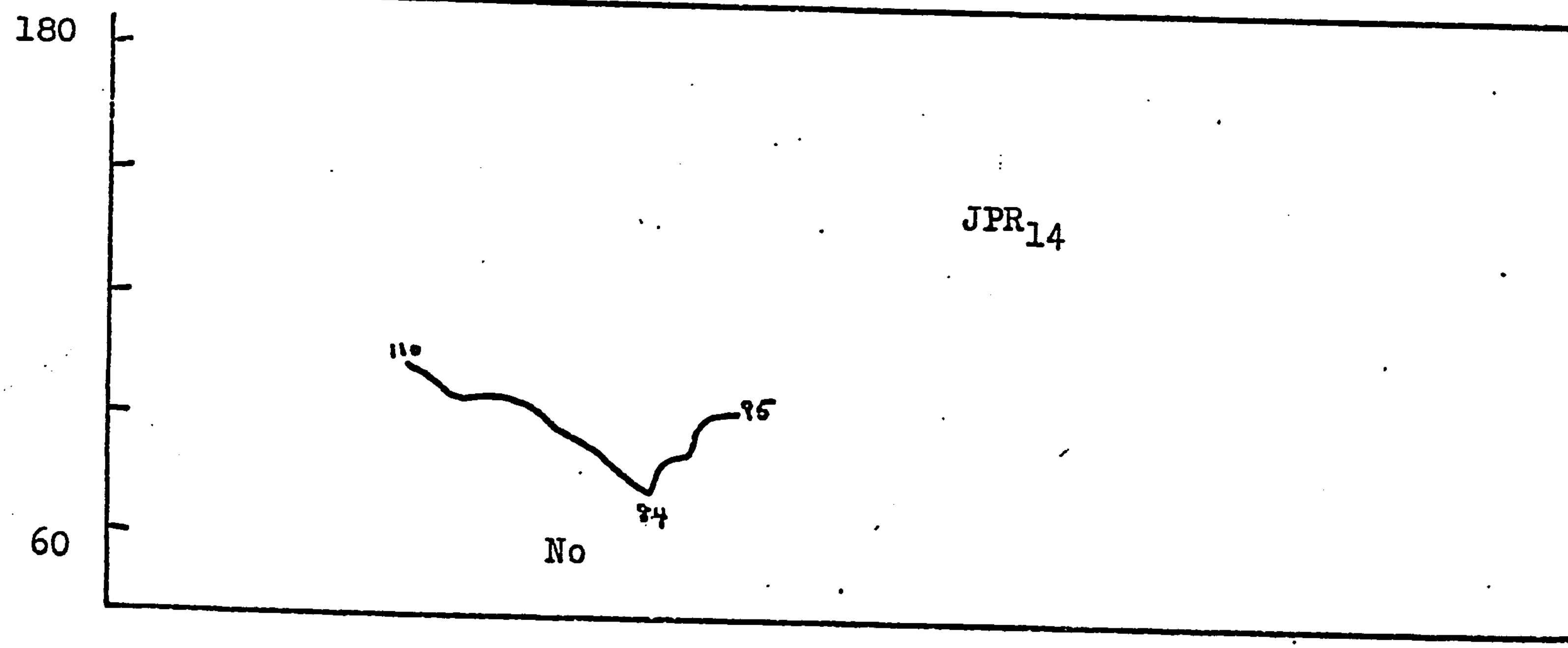

JPR7
240
60
107
144
99
111
118
82
101
108
115
But the slower kids can't do it and they don't want to do it

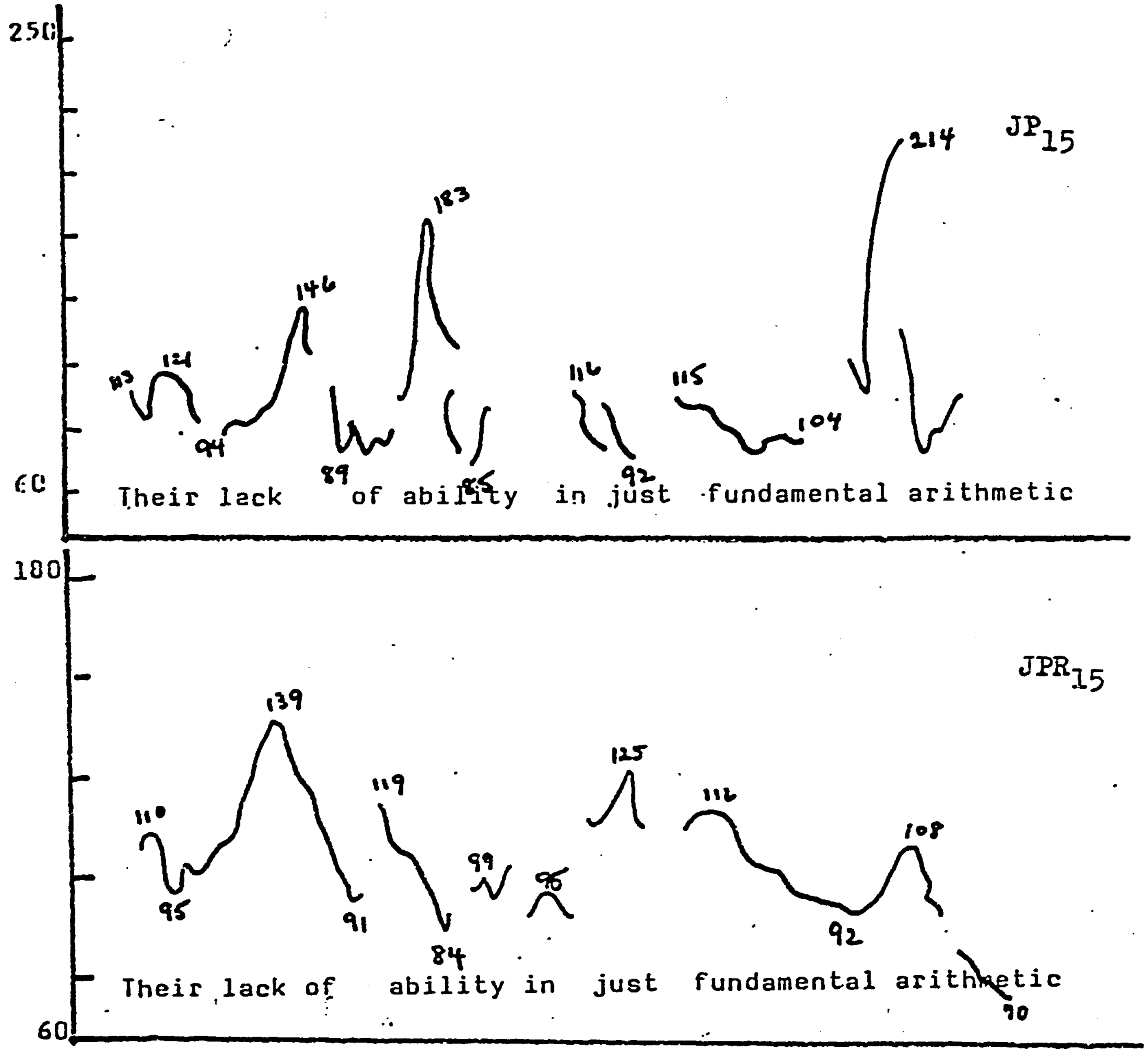


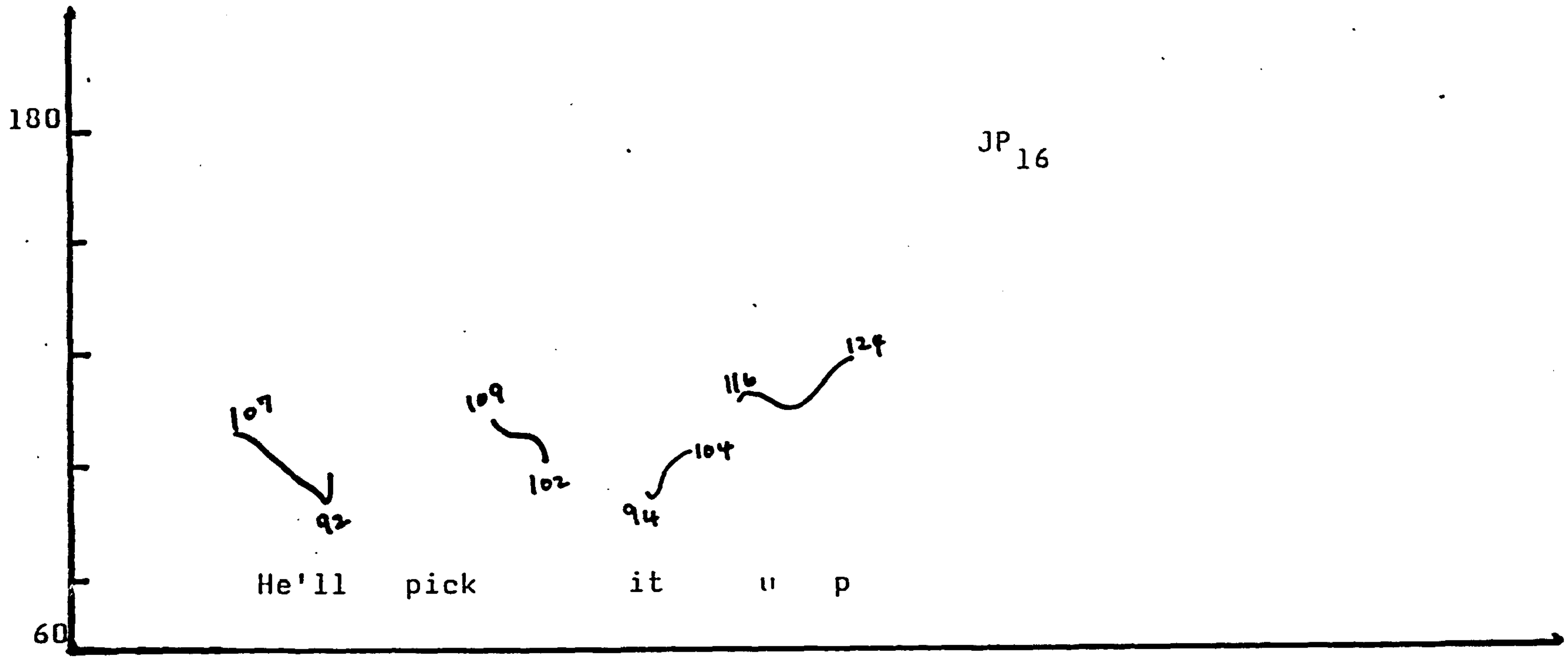
JP10
180
60
111
91
127
153
107
99
117
146
95
He's got a personality that doesn't quit.
JPR10
180
60
108
95
119
128
111
130
114
107
116
68
He's got a personality that doesn't quit.

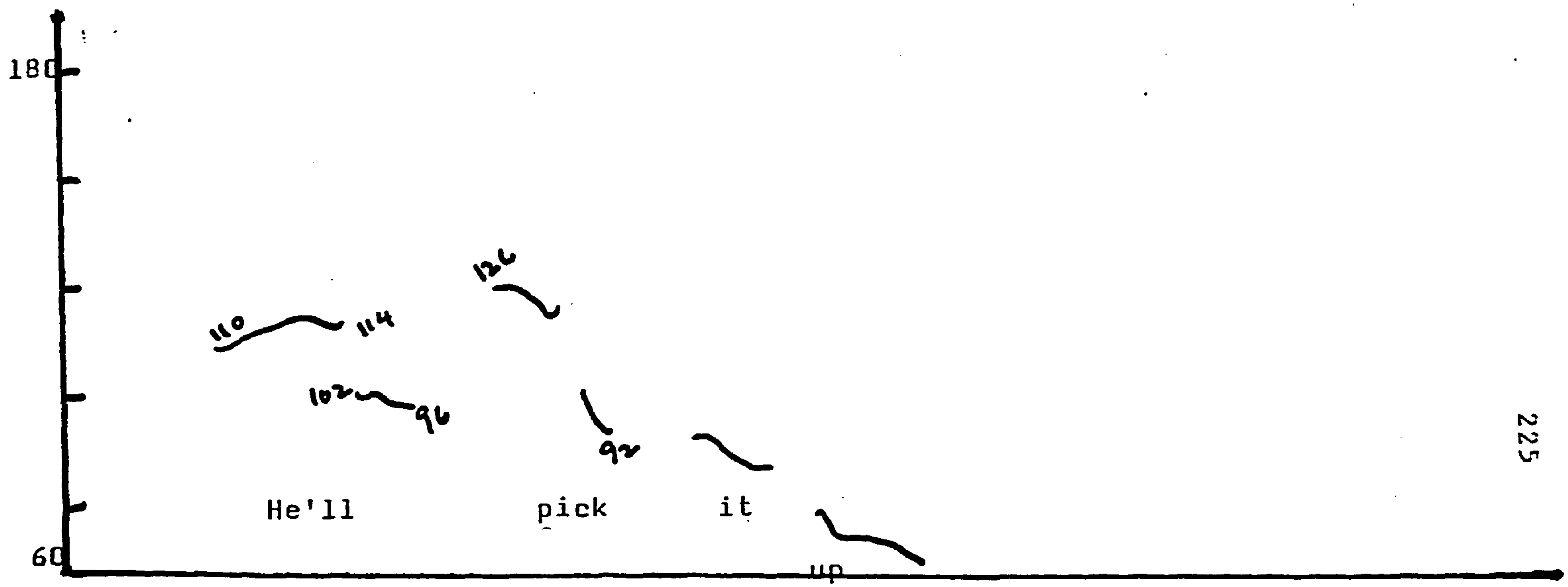

150
60
JP11
127
120
97
99
135
115
95
119
100
He can liven up any con ver sa tion.
150
60
JPR11
118
120
105
135
123
122
101
115
85
133
95
He can liven up any con ver sa tion.

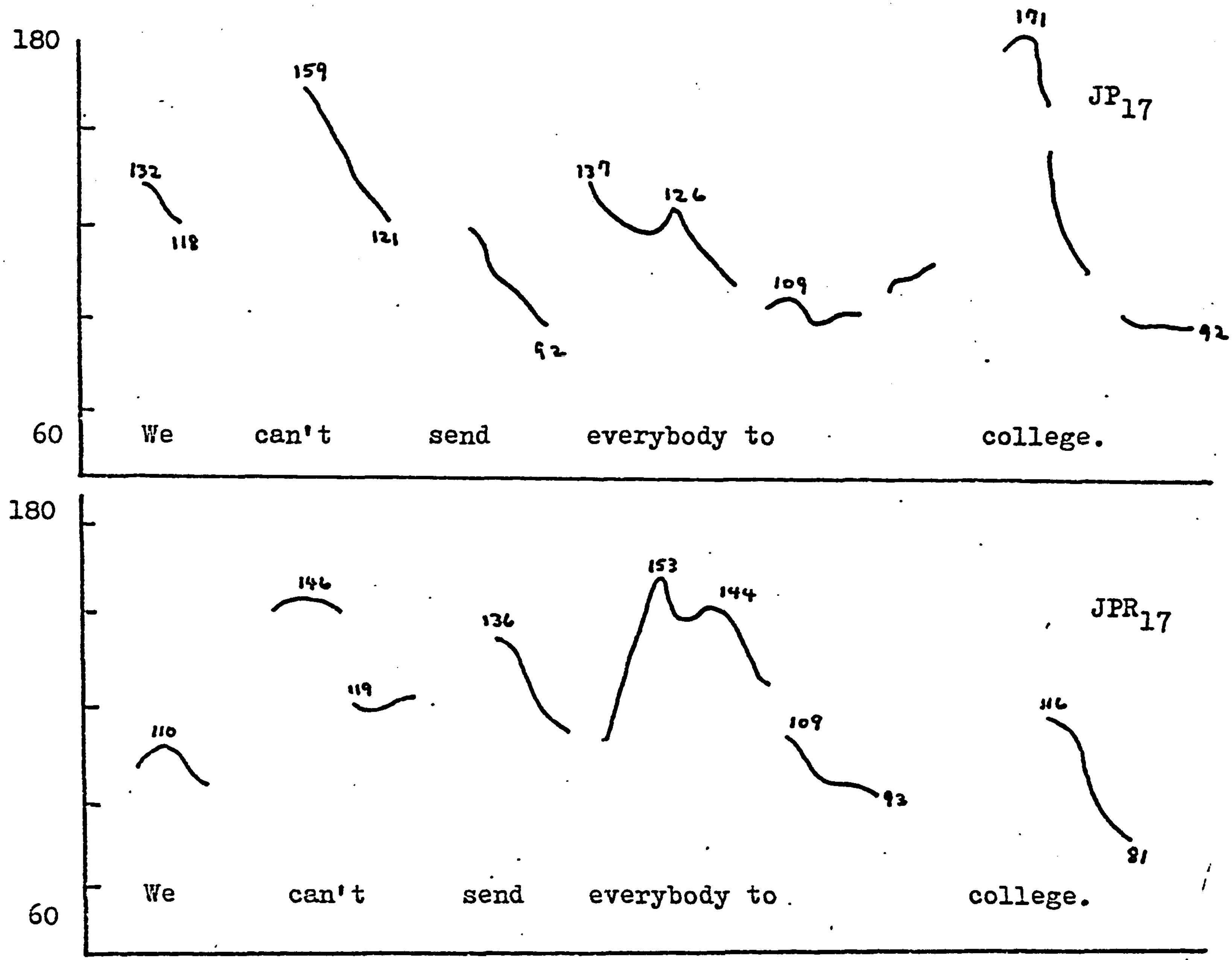

180
60
JP14
124
99
No


180
60
JPR14
110
84
95
No

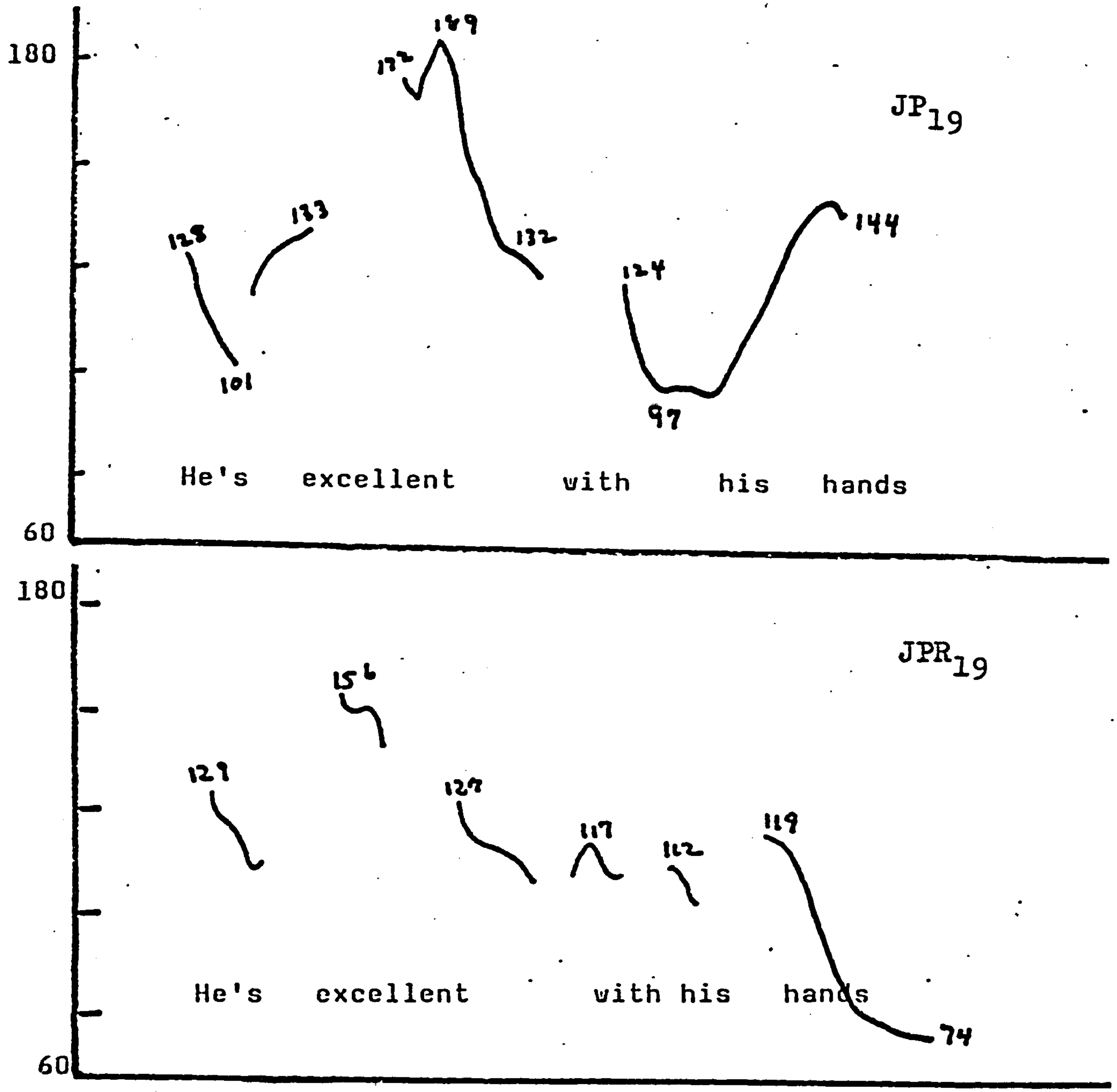

250
60
JP15
113
121
94
146
89
183
85
116
92
115
104
214
Their lack of ability in just fundamental arithmetic
180
60
JPR15
110
95
139
91
119
84
99
96
125
112
92
108
90
Their lack of ability in just fundamental arithmetic

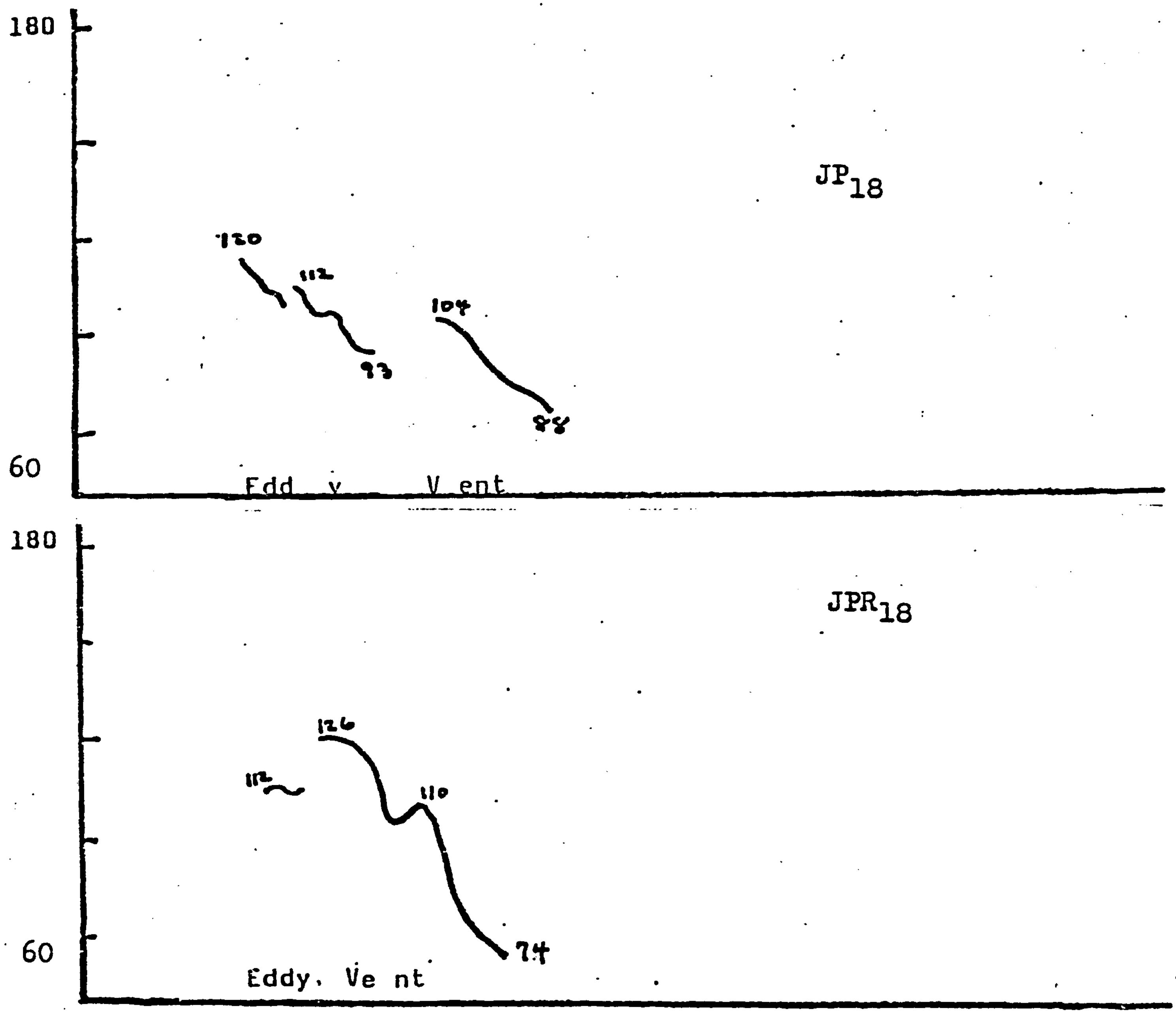

JP16
180
60
107
92
109
102
94
104
116
124
He'll pick it u p

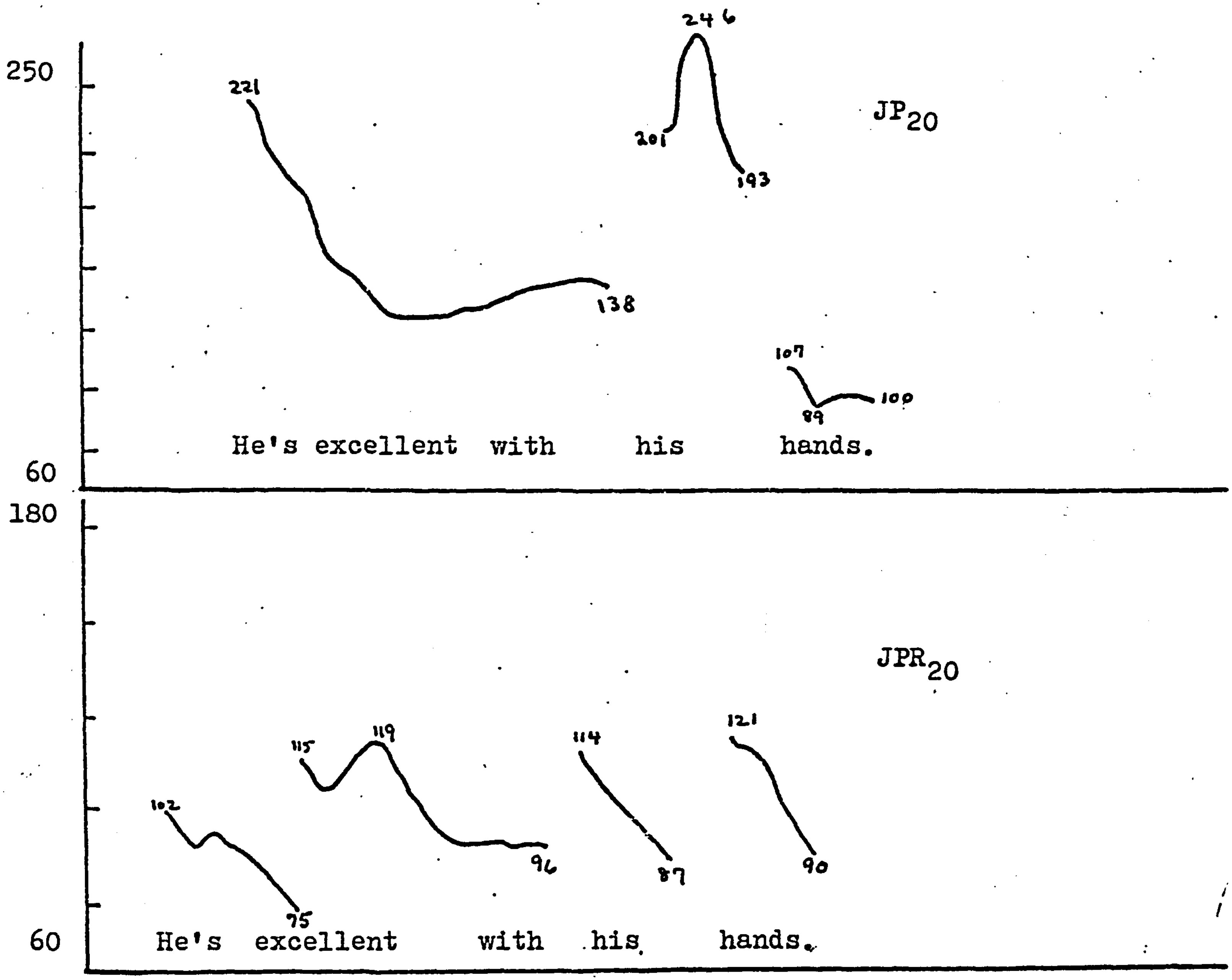

180
60
110
114
102
96
126
92
He'll pick it up

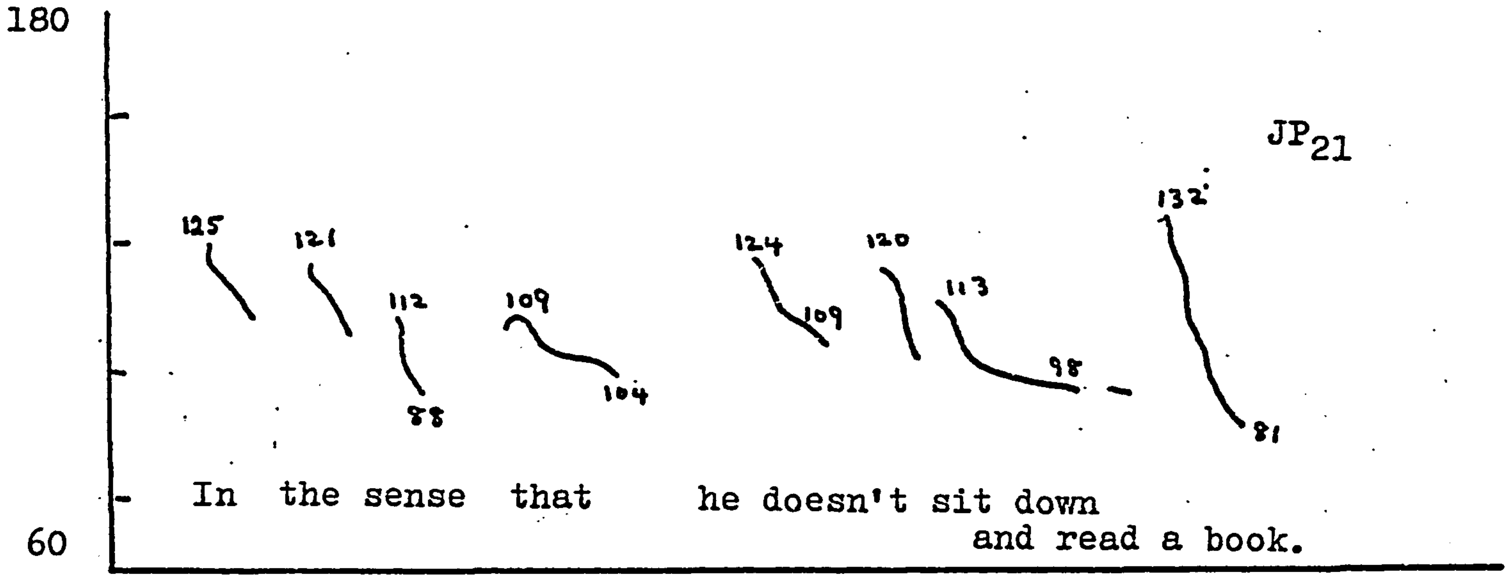


180
60
JP17
We can't send everybody to college.
132
118
159
121
92
137
126
109
171
92
180
60
JPR17
We can't send everybody to college.
110
146
119
136
153
144
109
93
116
81

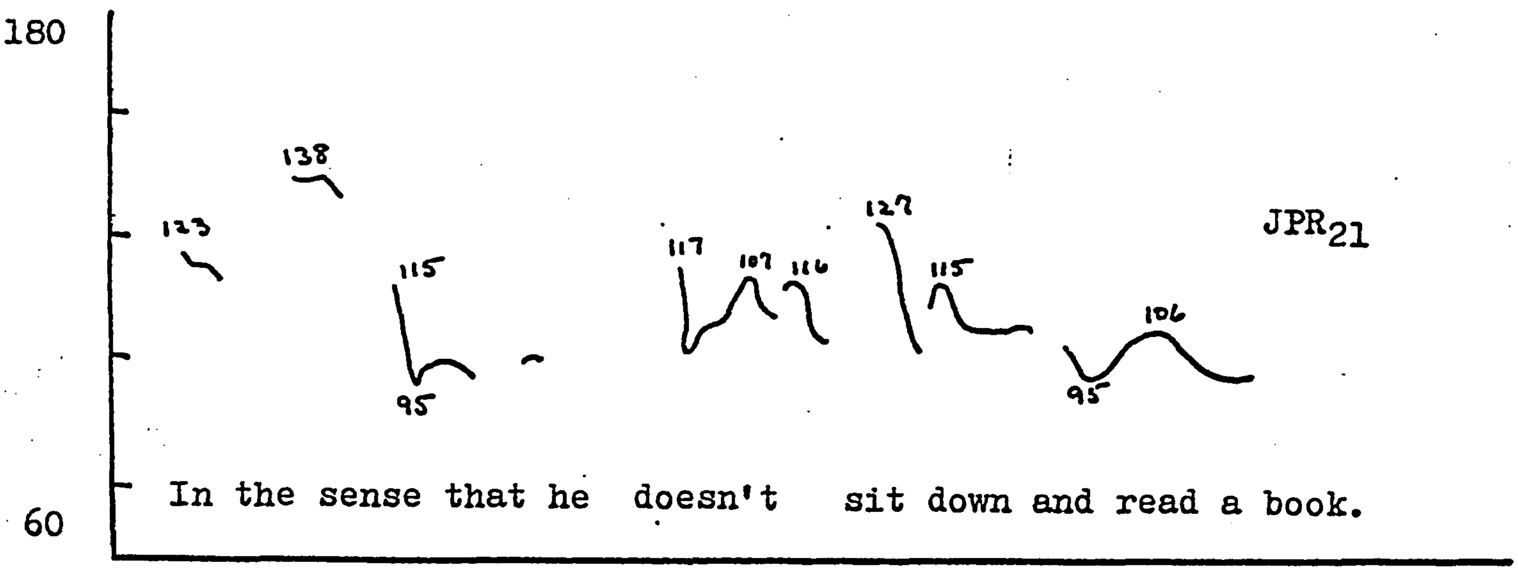

180
60
JP19
128
101
133
172
149
132
124
97
144
He's excellent with his hands
180
60
JPR19
129
156
127
117
112
119
74
He's excellent with his hands


180
60
JP18
120
112
93
104
88
Edd y V ent
180
JPR18
126
112
110
74
60
Eddy. Ve nt

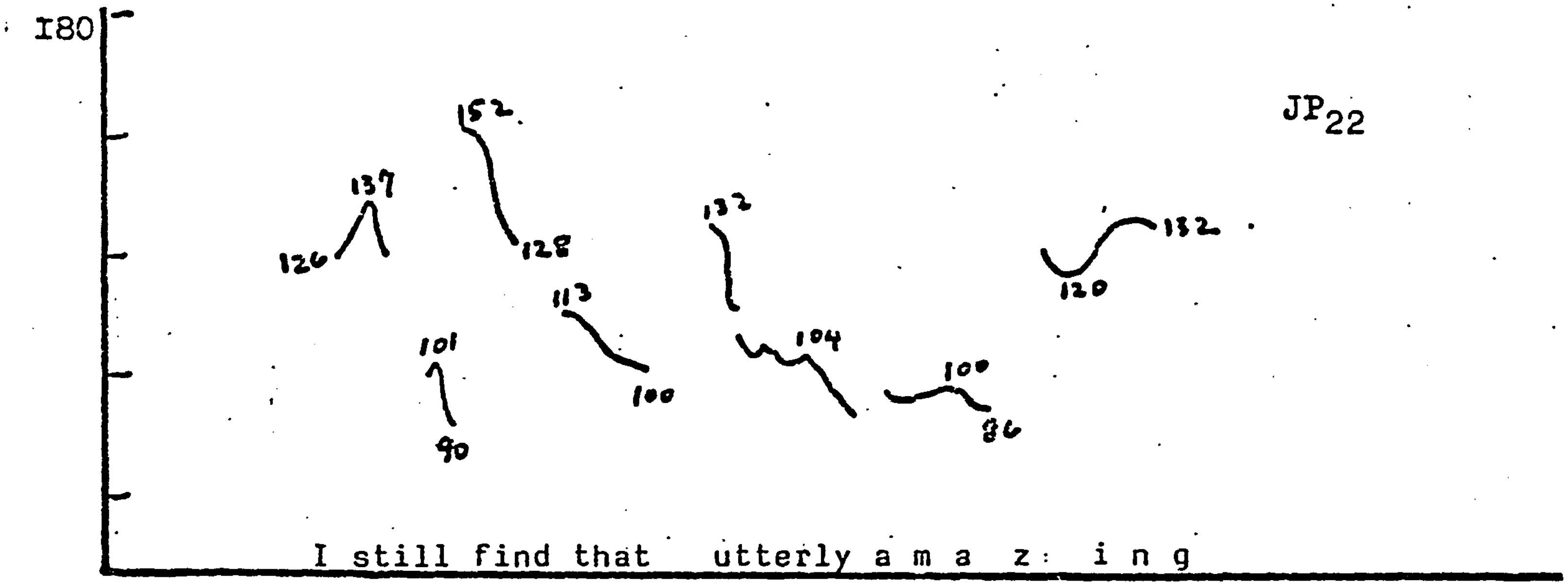

250
60
180
60
JP20
JPR20
221
138
201
246
193
107
89
100
He's excellent with his hands.
102
75
115
119
96
114
87
121
90
He's excellent with his hands.

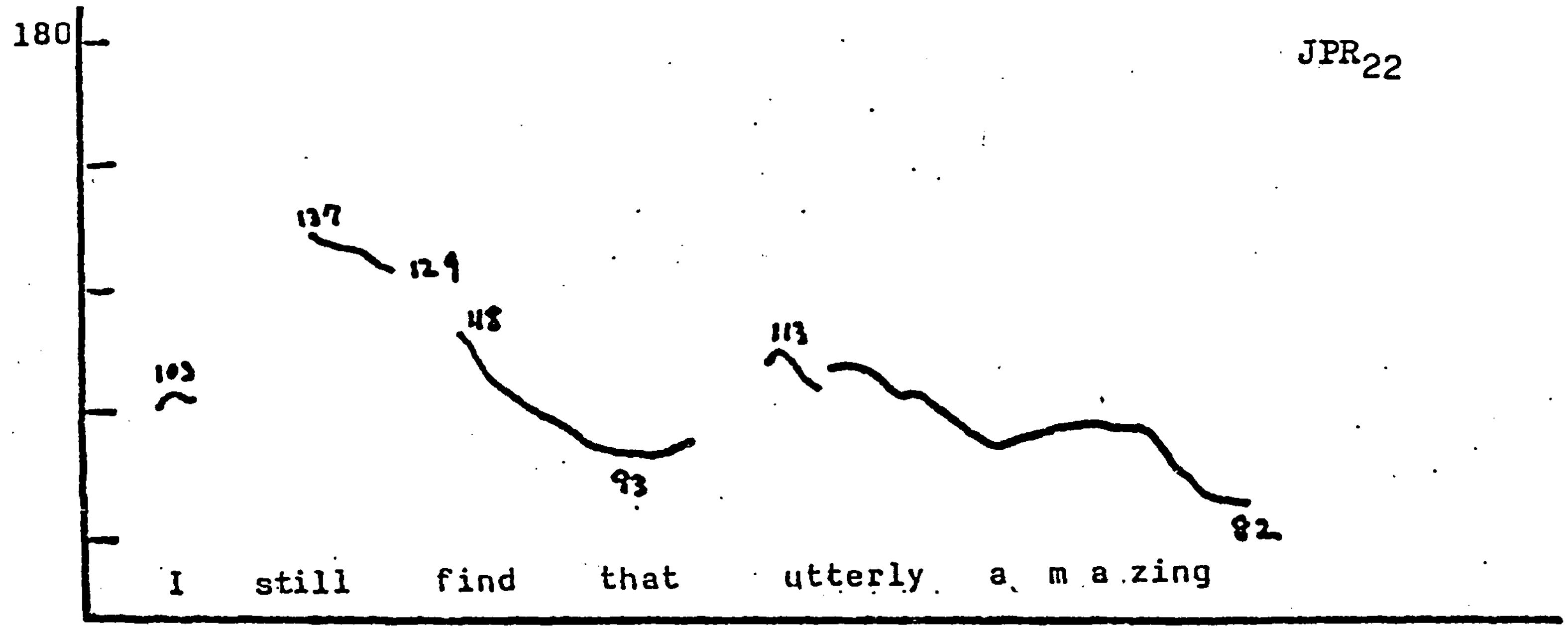


180
60
JP21
125
121
112
88
109
104
124
109
120
113
98
132
81
In the sense that he doesn't sit down and read a book.



180
60
JPR21
123
138
115
95
117
107
116
127
115
106
95
In the sense that he doesn't sit down and read a book.


180
JP22
137
126
152
128
101
90
113
100
132
104
100
86
120
132
I still find that utterly a m a z: i n g


180
JPR22
137
129
118
103
93
113
82
I still find that utterly a m a zing

REFERENCES

Abberton, E., and Fourcin, A.J. (1978). Intonation and Speaker Identification. Language and Speech., 21, 305-318.

Adams, C.M. (1969). A survey of Australian English Intonation. Phonetica, 20, 102.

Armstrong, L.E. and Ward, I.C. (1926). Handbook of English Intonation. Cambridge, England: W. Heffer and Sons (2nd Edition).

Atkinson, J.E. (1973) Aspects of Intonation in Speech: Implications from an experimental study of Fundamental Frequency. Ph.D. Thesis, The University of Connecticut.

Atkinson, J.E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. Jour. Acoust. Soc. Amer., 63, 211-222.

Baer, T. (1975). Investigation of Phonation Using Excised Larynxes. Ph.D. Thesis, Massachusetts Institute of Technology.

Barnwell, T.P. (1971). An algorithm for segment duration in reading machine contexts. Technical Report 479. Cambridge, Mass.: MIT Research Laboratory of Electronics.

Berman, A., and Szamos, M. (1972). Observations on sentential stress. Language, 48, 304-325.

Bloomfield, L. (1933). Language. New York, Holt.

Blumenthal, A.L. (1970). Language and Psychology: Historical Aspects of Psycholinguistics. New York: Wiley.

Bolinger, D.L. (1958). A theory of pitch accent in English. Word, 14 (2-3), 109-149.

Bolinger, D.L. (1965). Pitch accent and sentence rhythm. Forms of English. Abe and Kanekiyo (eds). Cambridge: Harvard University Press.

Bolinger, D.L. (1972). Accent is predictable (if you are a mind-reader). Language, 48, 633-644.

Boomer, D.S. (1965). Hesitation and grammatical encoding. Language and Speech, 8, 148-158.

Borden, G.J. and Harris, K.S. (1980). Speech Science Primer. Baltimore: Williams and Wilkins.

Breckenridge, J. (1978). Declination is a phonological process. Bell Laboratories Technical Memorandum.

Breckenridge, J. and Liberman, M. (1977). The declination effect in perception. Bell Laboratories Technical Memorandum.

Brend, R. (1972). Male-female intonation patterns in English. International Congress of Phonetic Sciences, 7, 1971.

Bresnan, J.W. (1971). Sentence stress and syntactic transformations. Language, 47(2), 257-281.

Cartwright, L.R. and Lass, N.J. (1976). A psychophysical study of rate of continuous speech stimuli by means of direct magnitude estimation scaling. Language and Speech, 18, 358-365.

Chomsky, N. (1971). Deep structure, surface structure and semantic interpretation. Semantics. Danny Steinberg and Leon Jakobowits, eds.

Chomsky, N. and Halle, M. (1968). The Sound Pattern of English. New York: Harper and Row.

Cohen, A. and t'Hart, J. (1967). On the antomy of intonation. Lingua, 19, 177-192.

Coker, C.H. and Umeda, N. (1971). Toward a theory of stress and prosody in American English. International Congress on Acoustics, 7.

Coker, C.H., Umeda, N. and Browman, C.P. (1973). Automatic synthesis from ordinary English text. IEEE Audio and Electro Acoustics, AU-21, 293-297.

Collier, R. (1975). Physiological Correlates of Intonation Patterns. Jour. Acoust. Soc. Amer., 58, 249-255.

Collier, R. and t'Hart, J. (1972). Perceptual Experiments on Dutch Intonation. Proceedings of the VIIth International Congress of Phonetics Sciences, (Montreal, August 1971), 880-884.

Cooper, W.E. and Sorenson, J.M. (1981). Fundamental Frequency in Sentence Production. New York: Springer-Verlag.

Crystal, D. (1969). Prosodic Systems and Intonation in English. London: Cambridge University Press.

Danes, F. (1960). Sentence intonation from a functional point of view. Word, 16.

Dittman, A.T. and Llewellyn, L.A. (1967). The phonemic clause as a unit of speech decoding. Jour. Pers. Soc. Psychol. 6, 341.

Erickson, D. and Atkinson, J. (1976). The function of the strap muscles in speech. Haskins Laboratories Status Report, 45/46.

Faaborg-Andersen, K. (1957). Electro-myographic investigation of intrinsic laryngeal muscles in humans. Acta physiolog. Scand, 41 Suppl. 140.

Fairbanks, G. (1940). Recent investigations of vocal pitch in speech. Jour. Acoust. Soc. Amer., 11, 457-466.

Fischer-Jorgensen, E. (1948). Sound duration and place of articulation. Zeitschrift für Sprachwissenschaft und Kommunikationsforschung, 17: 175-207.

Fry, D.B. (1958). Experiments in the perception of stress. Language and Speech, 1, 126-152.

Gaitenby, J.J. (1965). The Elastic Word: two experiments on the acoustic nature of stress. Haskins Laboratories Final Report, 9, Appendix D.

Garding, E. and Abramson, A. (1965). A study of the perception of some American English intonation contours. Studia Linguistica, 19.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. Jour. Acoust. Soc. Amer., 63, 223-230.

Gilbert, J.H. and Burk, K.W. (1969). Rate alterations in oral reading. Language and Speech, 12, 192-201.

Gleason, H.A. (1961). An Introduction to Descriptive Linguistics. New York: Holt, Rinehart and Winston.

Goldman-Eisler, F. (1956). The determinants of the rate of speech output and their mutual relations. Jour. Psychosom. Res. 1, 137-143.

Goldman-Eisler, F. (1961a). A comparative study of two hesitation phenomena. Language and Speech, 4, 18.

Goldman-Eisler, F. (1961b). The significance of changes in the rate of articulation. Language and Speech, 4, 171-174.

Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in Spontaneous Speech. London and New York: Academic Press.

Goldman-Eisler, F. (1972). Pauses, clauses and sentences. Language and Speech, 15, 114-121.

Grosjean, F. and Lane, H. (1976). How the listener integrates the components of speaking rate. Jour. Exp. Psych. 2-4, 538-543.

Hadding-Koch, K. and Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. Phonetica, 11, 175-185.

Haggard, M., Ambler, S., and Callow, M. (1970). Pitch as a voicing cue. Jour. Acoust. Soc. Amer., 47, 613-617.

Halle, M. and Keyser, S.J. (1971). English Stress. New York: Harper and Row.

Halliday, M.A.K. (1967). Intonation and Grammar in British English. The Hague: Mouton.

Halliday, M.A.K. (1970). Functional diversity in language as seen from a consideration of modality and mood in English. Foundations of Languages, 6, 322-361.

Harris, K.S. (1978). Vowel duration and its underlying physiological mechanisms. Language and Speech, 21, 354-361.

Harris, M.S. and Umeda, N. (1974). Effect of speaking mode on temporal factors in speech. Jour. Acoust. Soc. Amer., 56, 1016-1018.

Hirst, D.J. (1974). Intonation and context. Linguistics, 141, 5-16.

House, A.S. (1961). On vowel duration in English. Jour. Acoust. Soc. Amer., 33, 1174-1178.

House, A.S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. Jour. Acoust. Soc. Amer., 25, 105-113.

Jassem, W. (1972). The question-phrase fall-rise in British English. In Valdman, Papers in Linguistics and Phonetics to the Memory of Pierre Delattre.

Katsuki, Y. (1950). The function of the phonatory muscles. Japanese Jour. Physiol., 1, 29-36.

Klatt, D.H. (1973). Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. Jour. Acoust. Soc. Amer., 53, 8-16.

Klatt, D.H. (1976). Linguistic uses of segmental duration in English. Jour. Acoust. Soc. Amer., 59, 1208-1221.

Koshikawa, T. and Sugimoto (1962). The informational rate of the pitch signal in speech. Speech Communication Seminar: Stockholm.

Labov, W. (1966). The Social Stratification of English in New York City. Washington, D.C.: Center for Applied Linguistics.

Lass, N.J. and Lutz, D.R. (1975). The consistency of temporal speech characteristics in a repetitive oral reading task. Language and Speech, 18, 227-235.

Lea, W.A. (1972). Intonational Cues to the Constituent Structure and Phonemics of Spoken English. Ph.D. Dissertation. Purdue University.

Lea, W.A. (1973). "Segmental and Suprasegmental Influences on Fundamental Frequency Contours," in L.M. Hyman (ed.) Consonant Types and Tone. Southern California Occasional Papers in Linguistics No. 1, University of Southern California, Los Angeles, 17-70.

Lee, W.R. (1956). English intonation: a new approach. Lingua, 5.

Lee, F. (1969). Reading machine: from text to speech. IEEE Trans. on Acoustics and Engineering, 17.

Lehiste, I. (1970). Suprasegmentals. Cambridge, Mass.: MIT Press.

Lehiste, I. (1971). The timing of utterances and linguistic boundaries. Jour. Acoust. Soc. Amer., 51, 1228-1234.

Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. Jour. Phonetics, 4, 113-117.

Lehiste, I. (1979). Sentence Boundaries and Paragraph Boundaries--Perceptual Evidence. Unpublished Paper.

Lehiste, I., Olive, J. and Streeter, L. (1976). The role of duration in disambiguating syntactically ambiguous sentences. Jour. Acoust. Soc. Amer., 60, 1199-1202.

Lehiste, I. and Peterson, G. (1961). Some basic considerations in the analysis of intonation. Jour. Acoust. Soc. Amer., 33, 419-425.

Liberman, M.Y. (1975). The Int o national System of English. Ph.D. Thesis, Massachusetts Institute of Technology.

Liberman, M.Y. (1977). Modeling of duration patterns in reiterant speech. Bell Laboratories Technical Memorandum.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. Jour. Acoust. Soc. Amer., 4, 451-454.

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists. Word, 21, 40-54.

Lieberman, P. (1967). Intonation, Perception and Language. Cambridge: MIT Press.

Lieberman, P. and Michaels, S.B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. Jour. Acoust. Soc. Amer., 34, 922-927.

Lieberman, P., Sawashima, O., Harris, K.S. and Gay, T. (1970). The articulatory implementation of breath group and prominence: cricothyroid muscular activity in intonation. Language, 46, 312-327.

Lightfoot, M. (1970). Accent and time in descriptive prosody. Word, 26.

Lindblom, B. (1963). Spectrographic study of vowel reduction. Jour. Acoust. Soc. Amer., 35, 1773-1781.

Lisker, L. (1974). On "explaining" vowel duration variation. Glossa, 2, 233-246.

Maclay, H. and Osgood, C.E. (1959). Hesitation phenomena in spontaneous English speech. Word, 15, 19-44.

Maeda, S. (1976). A Characterization of American English Intonation. Ph.D. Thesis. Massachusetts Institute of Technology.

Martin, J.G. (1970). On judging pauses in spontaneous speech. Jour. Verb Learning and Verbal Behavior, 9, 75-78.

Mercer, N.M. (1973). Frequency and availability in the encoding of spontaneous speech. Language and Speech, 19, 129-143.

Molter, M.O. (1974). Durational relations between connected text, list sentences and carrier sentences. Jour. Acoust. Soc. Amer., 55, 460 (A).

Morton, J. and Jassem, W. (1965). Acoustic correlates of stress. Language and Speech, 8, 159-187.

Nakatani, L.H. and Aston, C.H. (1981). Acoustic and linguistic factors in stress perception. Bell Laboratories Technical Memorandum, 78, 1228-18.

O'Connor and Arnold (1963). Intonation of Colloquial English. London: Longmans.

Ohala, J. (1970). Aspects of the control and production of speech. UCLA Working Papers in Phonetics, 15, 1-192.

Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English. Jour. Acoust. Soc. Amer., 54, 1235-1247.

Olive, J. (1975). Fundamental frequency rules for the synthesis of simple declarative English sentences. Jour. Acoust. Soc. Amer., 57, 476-482.

O'Malley, M.H., Kloker, D.R. and Dara-Abrams, B. (1973). Recovering parentheses from spoken algebraic expressions. IEEE Transactions on Audio and Electroacoustics, AU-21, 117-220.

O'Shaughnessy, D.O. (1976). Modelling Fundamental Frequency, and Its Relationship to Syntax, Semantics and Phonetics. Ph.D. Thesis. Massachusetts Institute of Technology.

Palmer, H.E. (1922). English Intonation. Cambridge, England: W. Heffer and Sons.

Parmenter, C.E. and Treviño, S.N. (1935). The length of the sounds of a Middle Westener. American Speech, 10, 129-133.

Peterson, G. and Lehiste, I. (1960). Duration of syllabic nuclei. Jour. Acoust. Soc. Amer., 32, 693-703.

Pierrehumbert, J. (1978). Declination is a phonological process. Bell Laboratories Technical Memorandum.

Pierrehumbert, J. (1979). The perception of F_0 declination. Jour. Acoust. Soc. Amer., 66, 363-369.

Pierrehumbert, J. (1980). The Phonology and Phonetics of English Intonation. Ph.D. Thesis. Massachusetts Institute of Technology.

Pierrehumbert, J. (1981). Synthesizing intonation. Jour. Acoust. Soc. Amer.

Pike, K.L. (1945). The Intonation of American English. Ann Arbor: University of Michigan Press.

Pollack, I. and Pickett, J.M. (1964). The intelligibility of excerpts from conversation. Language and Speech, 6, 165-171.

Quinn, A.M. and Umeda, N. (1978). Estimate of speaking rate from a short sentence. Bell Laboratories Technical Memorandum.

Raphael, L.J. (1972). Preceding vowel duration as a cue to the voicing characteristics of word-final consonants in English. Jour. Acoust. Soc. Amer., 51, 1296-1303.

Raphael, L.J., Dorman, M.F. and Liberman, A.M. (1975). The perception of vowel duration in VC and CVC syllables. Haskins Laboratories Status Report on Speech Research, SR-42/43, 277-284.

Sapir, E. (1921). Language: An Introduction to the Study of Speech. New York: Harcourt, Brace and World.

Sawashima, M., Kakita, Y. and Hiki, S. (1973). Activity of the extrinsic laryngeal muscles in relation to Japanese word accent. Res. Inst. Log. Phon. V. Tokyo Ann. Bull., 7, 19-25.

Schmerling, S. (1974). A re-examination of "normal stress," Language, 50.

Shockey, L. (1973). Phonetic and Phonological Properties of Connected Speech. Ph.D. Dissertation. Ohio State University.

Smith, C. (1978). Differences Between Read and Spontaneous Speech of Deaf Children. Unpublished paper.

Stockwell, R. (1971). The role of intonation: reconsiderations and other considerations. UCLA Working Papers in Phonetics, 21.

Streeter, L.A. (1978). Acoustic determinants of phrase boundary perception. Jour. Acoust. Soc. Amer., 64, 1582- .

Suci, G.J. (1967). The validity of pause as an index of units in language. Jour. Verb. Learn. Verb. Behav., 6, 26.

't Hart and Cohen, A. (1973). Intonation by rule: a perceptual quest. Jour. of Phonetics, 1, 309-327.

Trager, G.L. and Smith, H.L. (1951). Outline of English Structure. Studies in Linguistics No. 3. Oklahoma: Battenburg Press.

Umeda, N. (1975). Vowel duration in American English. Jour. Acoust. Soc. Amer., 58, 433-445.

Umeda, N. (1978). Some notes on reading: talkers, material and reading rate. Bell Laboratories Technical Memorandum.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. Jour. Acoust. Soc. Amer., 70, 350-355.

Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. Language and Speech, 3, 223-234.

Van den Berg, J.W. (1958). Myoelastic-aerodynamic theory of voice production. Jour. Speech Hear. Res., 1, 227-244.

Wang, W. (1972). The many uses of F_0. In Valdman, <u>Papers in Linguistics and Phonetics to the Memory of Pierre Delattre</u>, Mouton.

Wells, R.S. (1945). The pitch phonemes of English. <u>Language</u>, 21, 27-39.

Wingfield, A. and Klein, J.F. (1971). Syntactic structure and acoustic pattern in speech perception. <u>Perception and Psychophysics</u>, 9, 23-25.

Zwicky, Arnold (1972). Note on a Phonological Hierarchy in English. <u>In Stockwell and Macauley, Linguistic Change and Generative Theory</u>. Indiana University Press.