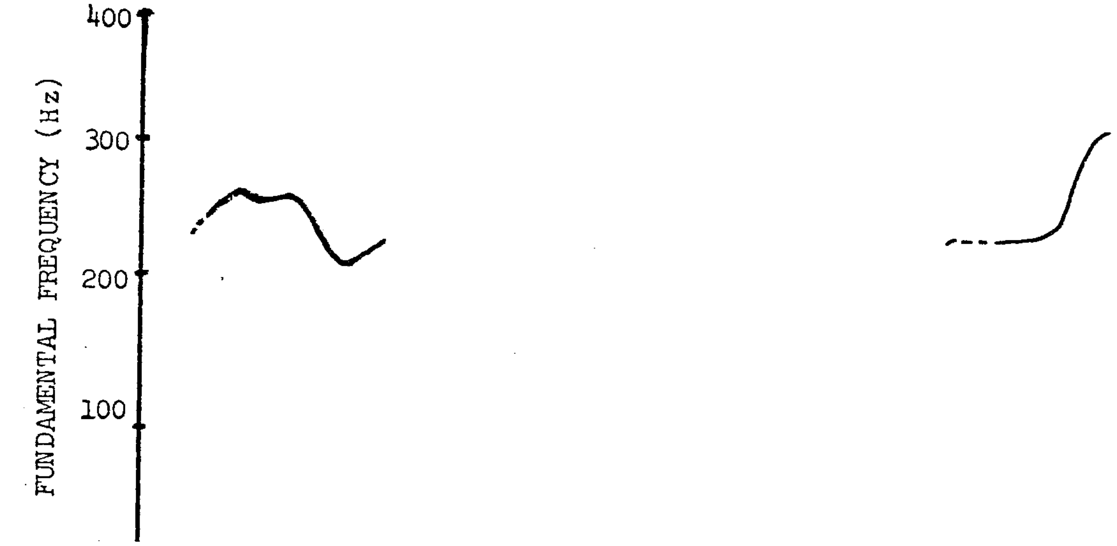
AN EXPERIMENTAL STUDY OF CERTAIN INTONATION CONTRASTS IN AMERICAN ENGLISH



S. ROBERT GREENBERG

WORKING PAPERS IN PHONETICS NO. 13
UNIVERSITY OF CALIFORNIA, LOS ANGELES
SEPTEMBER, 1969

UCLA

Working Papers in Phonetics

13

September, 1969

Extra copies are available from Textbook Department,
Student Store, UCLA, Los Angeles, California, 90024.
Price $2.50

An Experimental Study of Certain Intonation Contrasts in American English

by

S. Robert Greenberg

"Intonation is a half-tamed servant of language." -- Dwight L. Bolinger

"Obviously one can find out about competence only by studying performance, but this study must be carried out in devious and clever ways, if any serious result is to be obtained." -- Noam Chomsky

To my father,

Louis Greenberg

and my mother,

Rhea Gurian Greenberg

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ACKNOWLEDGEMENTS

My primary debt is to Professor Peter Ladefoged, who took a refugee from literary criticism and, with infinite patience, attempted to make an experimental scientist of him. In formulating and analyzing my experiment, I benefited from suggestions from Professors Ilse Lehiste of Ohio State University and Kerstin Hadding-Koch of the University of Lund, and from Professors R.P. Stockwell and George Allen of the University of California at Los Angeles. My understanding of intonation and language was enriched by discussions with Ralph Vanderslice, John Ohala, Tim Smith, Harry Whitaker, Mona Lindau and Raymond Silverstein of the UCLA Phonetics Laboratory. Mr. Silverstein also did the final proofreading assisted by Dale Terbeek. Stan Hubler, Willie Martin, John Lau and Larry Grant assisted me in the laboratory, and I am also greatly indebted to Mrs. Julie Haaker, Miss Jeanne Yamane, and especially Miss Michiko Yamane for typing the manuscript in the various stages of its gestation.

Part of this research was supported through ONR contract NR 049-226.

A Note on Terminology

In this study a strict delimitation is kept between the physical factors of fundamental frequency, amplitude, and duration of the speech signal, and their corresponding perceptual phenomena: pitch, loudness, and length. The term *stress* is used to refer to deliberately produced prominence, placed upon an English syllable through variations in any or all of the three parameters named above (plus, perhaps, variations in tempo and juncture), in which combination of factors variations in fundamental frequency capable of being perceived as pitch changes are taken as primary (cf. Chapter 2 for discussion of the primacy of pitch). Our definition of stress closely parallels Bolinger's use of the term *accent* (or pitch accent). We prefer to speak of *stress* for the following reasons: (1) We do not wish to be limited to Bolinger's conception of the number, shape, or functions of the various pitch accents in English. (2) We find it more convenient to contrast (sentence) *stress* with *word-stress*, than to use Bolinger's scheme of (sentence) *accent* versus (word) *stress*. Thus, when we speak of *contrastive stress,* we are always referring to sentence stress, not to a distortion of the normal stress pattern of a word (i.e. the inherent pattern predicted from the stress (cycle). (3) The nature of the stimuli used in our experimental study do not necessitate our making use of Bolinger's useful distinction between *contrastive* (sentence) *accent* and *contrastive* (word) *stress* (Bolinger, 1961c). In his example,

```
                port                   broad

I said to re
                   the trouble, not       cast it,
```

the words *report* and *broadcast* are contrastively accented, but normally stressed. In our own hypothetical example,

```
               re                        ex

I said to                        not          it,
                port the revolution,     port
```

the initial syllables of *report* and *export* show both contrastive accent and contrastive stress, in Bolinger's terminology. However, our data contain no examples of this type. The closest we come to a phenomenon of this type is the clearly gradient stretching of the first syllable in the "emphatic" enunciation of *ridiculous* (i.e. *r-i-i-dículous*). But this results not in a displacement of normal word-stress, but rather in the affective-semantic heightening of the prominence or stress of the entire word (cf. Chapter 4).

Chapter 1: Some Approaches to the Study of English Intonation

Problems in the Analysis of Intonation

The study of the intonation of English has a long, if not particularly distinguished, history. Pike (1945) in his survey of the field places the first important study in the year 1775, with the publication of Joshua Steele's *Prosodia Rationalis; or, An Essay Toward Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols.* Others might wish to consider John Walker, who in 1787 published *The Melody of Speaking Delineated; or, Elocution Taught Like Music; by Visible Signs,...* to be the author of the first worthy study of English intonation. Since that time, there have been hundreds of additional studies, and yet agreement exists (except among the authors of studies on intonation) that no language, not even English, has had its intonation yet described in a complete, accurate, and theoretically well-motivated way.

The problem is not one of the incompetence of linguists, but rather of the complexity of intonation itself. In any stretch of speech, several factors (some linguistic, some paralinguistic) operate simultaneously to yield the complex of perceptual elements which we call intonation. Uldall (1962: 783) has observed that in dealing with the intonation of real speech in real contexts "however much one may want to 'keep it clean', the fact is that the same kind of information is carried by several systems all present at all times: pitch, voice-quality, tempo, gesture, facial expression, any one of which, or combination of which, may be dominant at a given moment."

Even if, for reasons to be discussed in the next chapter, we accept the primacy of pitch as a cue for responding to an intonation contour, we are still not out of the quicksand. Daneš (1960: 40) points out the following strata which are relevant in hearing and interpreting pitch alone:

> (1) pitch as a component of the "stress complex"; (2) pitch in the quasi-phonemic system of intonation contours (functioning communicatively and otherwise); (3) the general pitch level of the rhythm-unit (utterance section) as a whole in relation to the neighboring rhythm-units (e.g., the low pitch level of parenthetic utterances); (4) the general trend of pitch in the whole utterance, or in longish parts of it (in compound rhythm-units with two or more intonation-centers); (5) the general pitch level of

> the utterance as a whole (in relation to the voice range of the speaker). The realization of elements belonging to a certain stratum depends on the realization of elements of all the strata lying higher up. [N.B. -- By "communicatively", I take Daneš to mean "syntactically" -- SRG]

The moment in which the linguist starts to interpret what he has heard is perhaps the most dangerous of all. As Bloomfield (1933: 114) observed, "Pitch is the acoustic feature where gesture-like variations, non-distinctive but socially effective, border most closely upon genuine linguistic distinctions." Is pitch merely affective? Or can it be used as a primary cue in intonation to mark syntactic distinctions? If so, can the syntactic and affective uses be clearly separated from each other, or is there a perhaps considerable area of overlap? Pike was well aware of the overlapping usages of intonation contours, and saw that the resulting complexity could lead the analyst into either of two extremes: bogging down in meaningles detail, or setting-up a grossly-oversimplified system. Seeing no clear-cut elements of syntactic intonation (e.g. a "question" pitch as distinct from a "statement" pitch), he chose to analyze intonation contours in terms of attitudes (1945: 24), some of which approach syntactic relevance, but which for the most part belong well within the affective domain.

Although it is still possible to find an occasional ghastly college speech text which recognizes only an emotional use for intonation in English, serious studies of English intonation during the past two decades have almost universally recognized a two-part division of intonation contours or "tunes", with statements, commands, and interrogative-word questions falling clearly into one category, "yes-no" questions into the other, and requests and similar locutions occupying a vague middle ground (the usual tendency being to assign requests to the "statement" category at first, then to seriously compromise this assignment in the later discussion). The statement/yes-no question division is made on the basis of contour end-shape: "falling" for statements, "rising" for yes-no questions. Only in rare cases do we find attempts to carry syntactic analysis of intonation contours significantly further (cf. Bolinger (1957), also Lees' critique of this approach (1960), and Bolinger's (1961) reply). Usually, the strictly syntactic analysis is gotten through as quickly as possible, sometimes in a few lines, and the writer launches into the discussion of "attitudes". Sometimes, as in Pike (1945), these attitudes subsume the syntactic element. More recently, as in Bolinger (1957), Hultzén (1957, 1964), less clearly in Trager and Smith (1951), the tendency has been to treat the syntactic element as basic, or "unmarked", and to consider the attitudinal (chiefly affective) elements as variations on, or overlays upon, this syntactic stratum.

However, all of these treatments share a common defect. Because the treatment of the syntactic features is so meager, and because the affective "vocabulary" of intonation is so rich, any *extended* treatment of English intonation (Jassem, 1952; Schubiger, 1958) invariably breaks down, as variation piles upon variation, "meanings" coalesce, and meaningful contrast dies. Reading some of these studies sadly reminds one of perusing the political diary of a madman, in which the clear universal principles of the first two pages, and the ringing exhortations of the last contrast pathetically with the gibberish and contradictions of the middle. And yet, in intonation studies as in politics, it is those who try hardest who fall farthest. Anyone can write a discursive two-page description of the intonation of English which does no violence to the essential facts. It is only when one tries to cover more and more intonational phenomena that one courts disaster.

It is our contention that a successful approach to the analysis of intonation must be based upon the following: (1) abandonment of the totally-unsupported assumption that all speakers within a particular geographical dialect area share exactly the same system of productive intonation contrasts; (2) utilization of a clearly-stratified model of intonation, in which certain functions are considered to belong to basic strata, and other functions are characterized as overlays upon these earlier (in the generative sense), more basic strata; (3) reasonably explicit hypotheses for connecting the syntactic element of the intonation system with the syntactic component of the grammar, making possible rule schemata for specifying the features of a particular intonation function (and specifying the stage at which it is inserted in the generation of an utterance), for specifying the feature-shape of alternate articulatory gestures (and the conditions under which they can be used), and for typifying rule changes, both diachronically and, in the synchronic plane, cross-dialectically; (4) a refusal to consider intonation models or rule schemata which do not receive support from a large amount of firm data, obtained from formal (i.e. fully controlled, specified, and replicable) experiments in which a strong correlation was obtained between certain physical features (i.e. contour shape, junctural phenomena) and certain psycho-acoustic or semantic responses.

Concerning the brave program sketched above, the first point is the heart of the present study, and will be discussed at length in Chapters 4, 5 and 6. The third point must be regarded as a project more for the future than the present. Some rather speculative suggestions will be made in Chapter 6, but they will not satisfy our own standards for explicitness. Point four has been adhered to, and it is hoped that the data from the experiment reported in this study will be of use to others working on studies of intonation. This leaves the second point. We present here a stratified model of English intonation, based largely upon the work of Hultzén and Bolinger. We shall introduce this model in Chapter 2, and modify it in Chapter 6. First, however, we must deal with the levels versus configurations controversy, and then assess some experimental studies of intonation.

Levels versus Configurations

As various students of intonation have pointed out (cf. Sledd, 1955), preference for an analysis of English intonation stated in terms of levels or in terms of configurations (i.e. "tunes" or "tone-patterns") is as much a matter of geography as of linguistic theory. With few exceptions (but cf. Bolinger, 1951), American linguists have tended to work in terms of levels; their British counterparts have preferred contours. Overlooking considerations of parochialness, there are reasons for the preferences. On the one hand, American linguistic theory of the 'thirties and 'forties (particularly the influence of Bloomfield) predisposed Americans to look for a phonemic or quasi-phonemic set of relationships in intonational phenomena. This approach culminated in the work of Harris (1944), Trager and Smith (1951) and Hockett (1955), in all of which operations previously employed in segmental analysis were carried-over into or adapted for the analysis of prosodic features. These linguists found attractive a system in which intonation could be specified in a small number of apparently discrete pitch levels, neatly tied to a system of stress levels and (later) junctures.

The British, on the other hand, built upon the musical-notation tradition and the moving-pitch tradition, which were respectively embodied in the eighteenth century by Steele and Walker, and at the turn of the twentieth century by Jones (1909) and Sweet (1892). The subsequent refinements into the "tunes" of Armstrong and Ward (1926), which Jones adopted from 1932 onward, and the segmentation of tunes into component parts such as "head", "nucleus", "tail", etc. put forward by Palmer (1922), and extended by such workers as Kingdon (1958) and Schubiger (1958), and even the "tone" and "tone-pattern" analyses of the neo-Firthian school, as in the works of Lee (1956), Sharp (1958), and Halliday (1963a and b) can all be seen as variations on a theme, rather than death and transfiguration of a theory.

There is at least one more factor (excluding, for the moment, "God's truth") operating in these preferences. British English shows a typical intonation contour beginning rather higher than the American contour, frequently sloping down to the nuclear section of the contour, and also frequently showing more variation of pitch in the pre-contour section. In other words, there is more "action" in the British pre-contour. Although we see the effect of such pre-contour variations as being almost entirely within the affective domain, it is nonetheless understandable that British linguists would prefer a kind of analysis which would be free of the pitch-point restrictions (i.e. restrictions stating that pitch need not be specified at more than, say, three points within an intonation contour) of a Trager-Smith level analysis, and would allow them to highlight various sections of the prosodic contour. In this respect it is interesting to note that it was a pre-contour phenomenon (Sledd's example of ^{2}cér^{3}tainly^1 #) which caused the participants in the 1957 Texas Conference to posit an increased number of (optional) pitch points in the intonation contour, and led to the system's becoming topheavy.

We come now to the central question: is an analysis in terms of levels or of configurations preferable for the analysis of English intonation, particularly of American English? The first point which must be made is that the two systems are largely convertible. For example, an American linguist working within the notational framework of numbered pitch levels will have little difficulty converting the configurational descriptions of a British linguist into his own kind of notation. In those cases where some difficulty does occur, the trouble frequently results from the fact that the British linguist is describing a pitch contour unfamiliar to American ears. Secondly, the dichotomization of intonation systems into level-analysis versus configurational-analysis is at least partly inaccurate. On the one hand, both Pike, and Trager and Smith made it clear that they were specifying intonation contours, not mere random collections of pitch-level sequences. In the Trager-Smith system, the pitch "phonemes" are explicitly assembled into intonational "morphemes" of the type $\sqrt{231\#}$. On the other hand, the British configurational analyst must, at least implicitly, have some idea of the range over which his tones move, and of what tonal range would differentiate, say, a low-rising tone from a high-rising one. In other words, one cannot draw a curve without assuming points (potentially representing levels) along the way. Thus, any contour assumes the presence of a set of levels, and any specification of pitch-pattern in terms of levels must (if it is to be adequate) operate within an assumed configurational framework. Ladefoged, looking toward a system of rules capable of activating a terminal-analogue speech synthesizer, also discards the strict dichotomization of levels versus configurations:

> In fact it seems clear that from the point of view of the higher level phonological rules, the complete contours contrast with one another; but the phonetic specification must be in terms of target pitches ... . The relation between intonation contours and target pitch levels is in some ways (but not in all ways) analogous to that between phonemes and the bundles of distinctive features or simultaneous categories of which they are composed. (Ladefoged, 1967: 52)

More than a decade ago, Sledd also saw that contour-analyses include the concept of levels, and he concluded that "Bolinger's antithesis between levels and configurations is ultimately false." (Sledd, 1955: 328; cf. also Hadding-Koch, 1961: 44-45)

But if the preceding discussion has somewhat cleared the air, it has not (nor has it attempted to have) eliminated all points of contention between these two schools of analysis. The remaining problems seem to us, however, to be a matter less of theory than of metatheory: specifically, those implicit assumptions underlying the theories of intonation analysis. One of these matters has received considerable attention during the past few years: the conflicting assumptions of the two schools concerning the relative independence of stress and pitch, with the level-analysts preferring to specify the two systems independently, and the configurationalists choosing to represent intonation contours as pitch curves into which stress phenomena were implicitly incorporated. As late as 1955, Sledd could mutter that "Palmer and Blandford seem to confuse stress and pitch ..."(1955: 328), but by the end of the 'fifties, the critics of level-analysis had won this particular point (cf. Chapter 2 for a discussion of the primacy of pitch). One of the last nails in this coffin was Lieberman's (1965) experiment, which showed that the Trager-Smith hypothesis of two independent four-level systems for stress and pitch led to a greatly-overspecified prosodic system. But the skepticism for which Lieberman supplied a measure of experimental verification had earlier been clearly expressed by Wang (1962), and earlier yet by Sledd, in his noted review of Trager and Smith:

> Though the *Outline* is much better barbered than (say) Pike's *Intonation*, its elegance is accompanied by a certain stiffness; and the system is severely tested by such phenomena as Bolinger's smoothly rising intonations, Pike's 'slurred precontours', or Pike's 'descending series of heavily stressed syllables' with more 'distinct pitches than can be fitted into four levels' (Pike, *Intonation* 67, 70). In the midst of a classroom exposition of Trager and Smith, there is considerable embarrassment in the sudden realization that in their notation the expositor could not write the intonations which he himself is using. (1955: 328)

However, Sledd immediately added, "The embarrassment would not be relieved by abandoning a levels-analysis for a contour-analysis like Palmer's and Blandford's," giving reasons which we summerized above.

Another crucial assumption which we have already mentioned concerns the number of pitch-points to be specified in a Trager-Smith type of intonation contour, the ruling assumption having been that the number must be very small (usually three) and invariable. There were two reasons for this view. First, it was felt that if there were not a small, specifiable number of pitch segments, then one could no longer predict the 2-levels, and one would have to write "2"s all across the sentence. (In this system, level 2 represents the normal, "carrier" frequency of the speaking voice.) Secondly, if one wished to integrate pitch "Morphemes" of this type with a generative grammar, then a small and specified number of pitch points would be necessary in order to be able to extract the elements from the sentence, and then put them back in place (cf. Stockwell, 1960).

However, this pitch-point requirement is much too strong. Concerning the first reason, there is increasing evidence that speech perception involves paying attention to certain kinds of pitch variations, and ignoring others, since the fundamental frequency (f_o) of the speaker's voice varies in a totally non-meaningful way as the supra-glottal articulatory mechanisms are adjusted for the articulation of different consonants and vowels. Indeed, if Flanagan (1968: C-1-6) is correct, the extent of these non-significant variations in f_o may easily reach 20 Hz, since in Flanagan's data the difference in f_o between /u/ and /a/ (at the same subglottal pressure and the same operating characteristics of the vocal cords) is as great as 40 Hz. It would then follow that level 2 would have to be considered to cover more territory than we might assume in the case of a mere "carrier" frequency. There is some evidence of this extensive range for level 2 for Swedish in Hadding-Koch (1961: 94ff.), where seven of the ten speakers are described as having a range of approximately 80 Hz for their level 2. Because of this wide frequency range for level 2, it would seem reasonable to assume a level 2 in any transcription, except in places where the transcription notes a different level.

One might dismiss the second, "generative", reason for strongly-specified pitch-points by simply stating that such considerations are rooted in a surface-structure oriented view of both syntax and phonology, a view which is now hopelessly outdated. We prefer, however, to discuss this reasoning further. One could, for example, claim that even within a deep-structure conception of a grammar, intonation assignment rules should be bunched well toward the end of the morphophonemic rules. But such a view inevitably contains a far too limited conception of the role of intonation within a grammar. Bierwisch (1966) has shown that intonation reaches deep into the grammar, and Lieberman (1967: 133ff.) has supplied an interesting hypothesis concerning relationships between deep structure markers and the realization of phonological features such as his "marked" breath-group (cf. our Chapter 2 and Chapter 6).

Perhaps even more serious is the fact that an overly-mechanical pitch-point specification ignores a tremendous amount of information available within the grammar from the inherent stress cycle. This information relates to what has been called "potential for pitch accent" (Bolinger, 1958: 137; Vanderslice, 1968: 53-54). To state the matter perhaps too simply, we do not need an overly-exact algorithm for, let us say, assigning nuclear stress to syllable X_i and not to syllable X_j next to it, because the inherent stress cycle of the grammar tells us that X_i is by far the more natural place for the assignment of that nuclear stress, so much so that we will assign nuclear stress to X_j only if there is present a special marker (of the general type EMPH, etc.) in the deep structure which alters our conception of normal sentence stress assignment.[8] Some interesting experimental verification for this view comes from Gårding and Gerstman (1960), who asked listeners to assign nuclear stress in the synthesized sentence "Where's he living now?" as the pitch-peak was moved from one place to another in the utterance. They report that "with movement of the intonation peak through the utterance there is, then, a regular progression of votes from *where's* to *liv* to *now*. When the peak is outside these three syllables it is generally referred to the nearest stressable syllable." (1960: 58)

In our own measurement of the intonation contours produced by the twelve speakers in our experiment, we utilized the four-point contour (pre-contour, peak, turning point and end point) found in Hadding-Koch and Studdert-Kennedy (1964). Our choice was made on purely empiric grounds, and in no way influenced the listening tests, which took place prior to the measuring of the stimuli.

We shall say very little about relativity of pitch levels, since we regard it as a pseudo-problem. Bierwisch (1966) uses as his starting-point Pike's statement that "The important feature is the relative height of a syllable in relation to the preceding or following syllables," and observes that this view makes possible two interpretations of pitch-transcription:

(a) a relative conception, but fixed to a specified zero-value;
(b) a relative conception, but with no fixed zero-value.

In such a system, *2 3* vs. *1 3* indicates only that the second differential is approximately twice the first, but tells us nothing about *how* great.

The first conception (a) makes possible a transcription of pitch relative to the pitch of the entire utterance. The second (b) makes possible only a transcription of pitch relative to neighboring pitch. Bierwisch prefers to work in terms of conception (b), because (1) Lieberman (1965) has shown that even in intelligently applied Trager-Smith notation "The pitch levels reflect the relative fundamental frequency only during segments of speech in which there is continuous voicing"; (2) Bierwisch states, "For the characterizing of linguistic conditions, it is utterly uninteresting how high a segment lies. Rather, it is important where pitch rises or falls lie, and whether these pitch discontinuities are great or less great." (1966: 135) To summarize: Bierwisch, wishing to concentrate upon "linguistic" (i.e. syntactic) intonation, chooses to ignore "gradient" phenomena of the type described by Bolinger (1961a). He succeeds in developing a workable set of generative rules describing a large part of the syntactic element in German intonation. Although these rules are not capable of serving as input to a speech synthesizer, they do show that relativity of pitch in itself is not destructive to significant work in the study of intonation systems. As Hadding-Koch observed, "Criticism would have been more justified ... had the level-analysts declared that listeners are able to hear *absolute* levels." (1961: 45)

We come now to the most serious question of all: *how many* levels (or, equally crucial, how many distinct *contours*) must be specified? This question was asked of level-analysts in the 'forties and early 'fifties, but has been a dormant issue for some years. It has never, we believe, been seriously raised in the discussion of configurational analyses. To the neo-Bloomfieldian Americans of the 'forties, such a question was an important one. Following essentially phonemic methods in their analyses of prosodic material, they wished to specify all and only the relevant items. Thus, Pike, defending his choice of a four-level specification, affirms:

> This number is not an arbitrary one. A description in terms of three levels could not distinguish many of the contours — for example, the three contours beginning on low pitch and each rising to a different height. A description in terms of five or six levels would leave many theoretically possible contrastive combinations of pitches unused. The four levels are enough to provide for the writing and distinguishing of all of the contours which have differences of meaning so far discovered, provided that additional symbols are used for stress, quantity, pause, general height of the voice, general quality of the voice, and so on. (1945: 26)

Similar reasoning resulted in four-level systems in the descriptions of Wells (1944) and Trager and Smith (1951). Hockett's system of three pitch-levels plus an additional "extra height" feature (1955: 45) differs only trivially from these.

Yet, as we have already noted (cf. page 6), an inflexible four-level system can easily be put under severe strain by various frequently-occurring pitch variations in intonation. Exactly the same situation holds for configuration-analyses, although this fact can be hidden in the general fuzziness of contour-descriptions. The situation might be likened to that in a legal contract: the more open and simple the initial terms, the longer and more dense will be the fine print. The two-tune analysis of Armstrong and Ward (1926) is followed by a great deal of "fine print", and in the case of Kingdon, the qualifications and variations, including the famous "thirty-six variations on Tone Five" (1958: 141ff.) are almost staggering.

No one has really undertaken to apply a simplicity metric to a level-analysis and a competing configuration-analysis of the same dialect of English (it being absurd to try to compare an analysis designed for Southern British with one designed for General American). Furthermore, early attempts at comparison (cf. Sledd, 1955), tended to concentrate upon the independent specification of pitch and stress in the Trager-Smith analysis as an essential element of the comparison. Since we consider both the standard level-analyses and the competing configuration-analyses to be basically wrong, in that they assume the same system of intonation-contrasts for *all* speakers of the same geographical dialect (cf. page 3), we shall not attempt such a comparison. However, we will suggest at least the following: (1) the implicit assumption underlying all configuration-analyses is the belief that disparate elements can be simply and accurately described by considering them as one unitary entity, in this case an intonation-contour; (2) additionally, there is the implicit assumption that, in a successful analysis of this type, the total number of discrete unitary entities will be quite small; (3) however, we believe that a configuration-analysis of English (British *or* American) begins to approach ideal coverage of the everyday data only as it begins to approach the complexity shown, for example, in Jassem (1952: 60), whose system is composed of twelve separate nuclear contours (eight unidirectional, four bidirectional), not including pre-contour variations. Confronted with such a number of basic entities, a linguist must begin to wonder whether some simplification is possible, one which will not require the kind of "fine print" referred to earlier. A possible simplification would be in terms of distinctive features, in the manner of Wang's analysis of word-tones (Wang, 1967). But such an approach necessarily assumes that all elements serve the same function (as the different word-tones all serve to discriminate lexical items), and we believe that the situation with sentence intonation is more complex than this.

In the next chapter, after surveying some basic physical and perceptual facts underlying the production and perception of intonation, we shall propose a model of intonation, one in which the different functions are assigned to different strata, one overlaid upon the other. In Chapter Six we shall extend this basic model by proposing ways in which different speakers (from the same geographical area) might have systems with different numbers of contrasts.

Chapter 2: The Experimental Study of Intonation Phenomena: "Primacy" and "Archetypality"

Experiments and Pseudo-experiments

Far too many dogmatic statements in phonology are the product of the "armchair" school of linguistic investigation. The dangers of this approach are great even in the investigation of segmental phenomena in a foreign language. If one is studying one's native language, the biases increase, at least partly because "linguist" and "native informant" tend to be two hats upon the same head. And when the subject under investigation consists of suprasegmental phenomena, the dangers are greater than ever, for three reasons: first, whether one is testing one's own *Sprachgefühl*, or running down the hall to test a colleague's, it is almost impossible when dealing with suprasegmentals, particularly on the spur of the moment, to find anything approaching a minimal pair test; secondly, it is difficult for even a trained listener to concentrate on or isolate a particular prosodic element submerged in a mass of other cues (we are generously assuming the possibility of unexaggerated enunciation), both segmental and suprasegmental; thirdly, even if the listener(s) should approve the prosodic variation which the investigator is "testing", this proves only that such a variation *could* be used in the language. It does *not* prove that it is usually used, or even frequently used (only neutral elicitation techniques employed with a number of speakers can give such information), nor does it prove that it is connected with any particular meaning (only rigorously-controlled tests of a large number of listeners can tell us this). The third point is especially relevant in the case of those serious students of intonation who claim some measure of verification for their analyses of English intonation from the fact that they have listened to X-number of hours of recorded telephone conversations, Shakespeare performances, faculty trysts, or whatever. All three objections apply in the case of the investigator who says something like, "I was wondering about the uses of these two contours, so one day I came into class and performed a little experiment..." A typical example of this sort of "experiment" can be found in Jassem (1952: 47-49), where the experimental procedure features such classic psychophysical blunders as the use of a single speaker who is well-known to the listeners, *live* presentation of speech samples by a speaker in full view of the listeners (who thus become listener-watchers), etc. However, the difficulty of designing carefully-controlled experiments on suprasegmentals makes us appreciate all the more those who have succeeded in so doing. We will now turn to some important studies concerning the production and perception of intonation.

The Primacy of Pitch

In analyzing any complex phenomenon, one must first attempt to isolate the component elements. Then, by using these elements as parameters capable of variation in series of psycho-acoustic tests, one can attempt to determine which elements are primary, which secondary in eliciting responses usually associated with the entire complex. If one succeeds in establishing the primacy of a particular element, then one can finally embark upon truly quantitative studies determing, for example, how great a quantitative change in the primary cue is necessary to cause a qualitative change in response to the complex.

Exactly this process has been undertaken in the study of prosodic features during the past fifteen years, and it is interesting to compare the before-and-after situation in phonological theory. We believe it fair to say that it was not only the members of the Trager-Smith school who believed in the primacy of stress (i.e., "intensity" of "utterance force"), and who believed that stress and pitch needed to be thought of as independent and equally significant elements. The dispute, as we see it, centered more upon the issue of whether one needed to specify these elements separately in a *notation*. It was thus largely an issue of notational economy, rather than a question of the fundamental perceptual nature of intonation.

Then, in the mid-'fifties, several groups of investigators began to isolate the relevant elements of intonation, and to test them in the way we have just described. Almost all of this testing was done on utterances of one-word length, tested either with natural or synthetic speech samples, the typical test contrasting pairs such as *dígest/digést, cónvict/convíct*, etc. The results were rather surprising. Fry (1958) found that fundamental frequency, intensity and duration were all relevant cues, but that "experiments with more complex patterns of fundamental frequency suggest that sentence intonation is an over-riding factor in determining the perception of stress and that in this sense the fundamental frequency cue may outweigh the duration cue" (1958: 151). Denes and Milton-Williams reported that "fundamental frequency, as well as intensity and duration, could be used by listeners to make intonation judgments, but that fundamental frequency provided the dominant cue in those intonation groups associated with large frequency changes" (1962: 11). Lieberman (1960) also showed the importance of f_o in the acoustic make-up of stressed syllables, as did Lehiste and Peterson (1961). Bolinger (1957-58a) proved that adding intensity to a low-intensity stress failed to improve the quality of the pitch accent, and Bolinger and Gerstman (1957) showed that some "stress" effects were really related to disjuncture phenomena.

Considering all this, it is not surprising to come upon Rigault's findings: On a single-factor experiment, varying f_o got a prominence rating of 70%, as against only 15% for intensity and 15% for duration (1962: 739). At the same time, psycho-acoustic testing of a more elemental nature showed why this should be the case. For example, Flanagan (1957), after warning that "the ability of man to make absolute discriminations is considerably less acute than his ability to make differential discriminations" nonetheless notes data which would pose severe problems for anyone wishing to posit a system in which signal intensity or amplitude would function as the primary cue. Specifically, he notes that the difference limens (DL) for formant frequency are of the order of ± 3% of the formant frequency and the DL for fundamental frequency is of the order of ± 0.5% to ± 1.0% for a vowel having a fundamental frequency in the neighborhood of 120 Hz. However, the DL for second formant *amplitude* is of the order of ± 3 db, or ± 40% of the formant amplitude, and the DL for over-all vowel amplitude is approximately ± 1 db, or about ± 12% of the over-all amplitude.

Early reports of these findings, and of other work later reported in Bolinger (1965: 17), enabled Bolinger to state that "the primary cue of what is usually termed STRESS in the utterance is pitch prominence ... . Intensity is found to be negligible both as a determinative and as a qualitative factor in stress ... [and that] while the upward obtrusion is basic, pitch prominence need not be merely upward, as commonly supposed, but may take other directions." (1958: 149) Such proofs of the primacy of the f_o parameter have made possible further experiments such as that of Gårding and Gerstman (1960), which we have already mentioned, as well as the work of Uldall (1960, 1962, 1964). They also underlie the experimental work reported in Chapters 3 and 4 of the present study.

On "Archetypes" in Intonation

Between the "how"of the empirical scientist and the "why" of the natural philosopher or theologian, there lies what we might call the "meta-how". What we have discussed in the preceding section constitutes a "how": whatever the meanings of intonation contours, they come about through variations of several physical speech parameters, of which fundamental frequency and the resulting perception of pitch seem to be primary. There is little grist here for the mill of the natural philosopher, let along the theologian. But if we were able to show, because of certain limitations or predilections in the physical make-up of human speakers and hearers, that the "how" could not be otherwise than it is, or (less satisfactorily) that the "otherwise" is clearly a distortion of or later superimposition upon our basic, necessary scheme, we would have our "meta-how", which could then be ridden off into the sunset of a discussion of "evolution", or of "God's Far-sighted Plan", or whatever. Owing to the increasing vulgarization of Jung's ideas by workers in other fields, it has lately become fashionable to call our "meta-how" by the resounding name of "archetype."

An "archetypal" theory of intonation has recently been proposed by Lieberman (1967), in a minor revision of his 1966 dissertation. As such theories go, it is of minor scope, applying in its strong form only to *American* English. However, frequent attempts are made throughout Lieberman's book to extend this theory to other varieties of English, as well as to other languages (Lieberman, 1967: 131-33). There are other reasons why this theory deserves attention: first, Lieberman's earlier work has been intelligently and rigorously conceived, and illuminating in its results; secondly, *any* archetypal theory is necessarily worthy of close attention. Because it states that certain bases could not be otherwise than they are (or are said to be by the theorist), such a theory, if accepted, inevitably places severe limitations upon our conception of such bases. If the theory is correct, then it advances science, by eliminating idle speculation. But if, as is too frequently the case, it is wrong, then it damages science by forestalling the "idle speculation" of a Galileo or Einstein. Therefore, any archetypal theory should be immediately subjected to attempts at confirmation or disconfirmation.

Lieberman's theory might be said to begin with the well-documented observation (Bolinger, 1964a) that an overwhelming percentage of the world's languages utilize a basically falling intonation contour for statements, and contrast with this falling contour a not-falling contour, especially in the case of certain types of questions, requests, and various other varieties of non-declaratives. Is there a reason for this tendency to exist? In stronger terms, *must* it be this way? Lieberman, in choosing to speak of "archetypal" patterns in intonation, obviously believes that it must, and wisely reasons that a physiological constraint or series of contraints would be both easier to prove and easier for the scientific community to accept than would, say, a psychological explanation (e.g., "finality" as an inherent human concept). However, the nature of physiological evidence is such that it becomes obvious when a hypothesis has not been proved. In the present case, not only has Lieberman not proved his hypothesis, but at several points his evidence argues directly against him.

The theory in question is based upon the belief that the fundamental frequency of phonation is largely a function of subglottal pressure, so that rises and falls in f_0 are directly related to rises and falls in the P_s (subglottal pressure) curve. However, in such a simple form the theory would obviously be false, since non-falling contours are manifested in cases such as English yes-no questions, despite a possible (Lieberman would say "necessary" or "archetypal") fall in subglottal pressure at the end of the utterance. To cover this situation Lieberman proposes a "marked-unmarked" distinction. In the case of the "archetypal unmarked breath-group" [*-BG*], the fundamental frequency of the vibrating vocal cords "appears to be a function of the subglottal air pressure and rises from a medium pitch to a higher pitch at the stress peak (which occurs at the peak subglottal air pressure) and then falls as the subglottal air pressure falls at the end of the utterance." (Lieberman, 1967: 27) Lieberman, in fact, specifies not merely a fall, but an "abrupt fall" in the P_s, occurring within the last 150-200 msec of phonation. Contrasting with this pattern is that of the "marked breath-group" [*+BG*], which manifests "an increase in the tension of the laryngeal muscles at the end of the breath-group where the air pressure falls" (53). Interacting with these two suprasegmental features is a third, segmental feature labelled "prominence" [$+P_s$], defined as "a momentary increase in the subglottal air pressure that is superimposed on the breath-group by the activity of the respiratory muscles. This momentary increase in subglottal air pressure can occur at any part of the breath-group except at the very end of the breath-group." (53-54)

Before turning to a detailed discussion of Lieberman's breath-group theory, let us note some problems in the formulation of his "prominence" feature. In the first place, there seems no good reason (aside from Lieberman's hypothesized physiological constraints) why prominence should not be capable of occurrence within the last 150-200 msec of phonation. One need only think of an exclamation such as "You're doing WHAT?" to realize that an EMPHASIS marker in the deep structure can be manifested on an item occurring at the end of the surface string, requiring a heavy degree of prominence (or "stress") at the end of the utterance. Furthermore, Lieberman provides us with such an example in his Figure 4.15, which gives the quantized spectrographic, f_0 and P_s records of Speaker 1 saying "Did Joe eat his *soup*?" We know, from two different kinds of evidence, that *soup* is an example of "prominence" in this sentence. First, it is italicized, and Vanderslice reminds us that in Lieberman's book

> ... the captions to figures 4.10-4.33 (his main data) note occurrences of [$+P_s$] in all and only the instances where the subject read an italicized word. Peaks of subglottal pressure appearing with other syllables, whether elsewhere in the same sentence -- higher peaks than for the italic syllable in 4.15 and 4.20 -- or in sentences with no italic indication of emphasis (4.18, 4.22, 4.23), are ignored. Thus the truly archetypal correlate of prominence seems to be not physiological, nor acoustic, nor perceptual, but typographic. (Vanderslice, 1969: 434)

Secondly, the rise on *soup* is much greater than that shown by the same speaker in Figure 4.13 ("Did Joe eat his soup?") or Figure 4.14 ("Did *Joe* eat his soup?"), thereby showing that this is indeed a case of prominence occurring during the last 150-200 msec of phonation. A further difficulty results from Lieberman's failure to prove that the f_0 rises on prominent syllables are largely a function of rises in the P_s curve. His inferences here are subject, therefore, to the same kinds of objections made below concerning his basic breath-group theory.

As we have stated, Lieberman seeks to prove the existence of physiological constraints upon intonation in human languages, and by means of those constraints to cast light upon some of the universal or quasi-universal attributes of intonation. It is immediately obvious that in arguing for the existence of such physiological constraints, he is also arguing that they are innate, and he is thereby forced to argue their presence at the earliest possible stage in human language activity, to argue their importance in adult language behavior, and to deny the importance of intonational behavior which does not show such restraints. He does this by designating the P_s curve with a final fall as "innate", by attempting to demonstrate that infant cries directly reflect that P_s curve, by arguing an "archetypal" status for the unmarked breath-group which allegedly resembles the "innate" P_s curve, by arguing that, no matter how an adult speaker produces a particular intonation curve, he *perceives* that curve as an expression of an archetypal pattern, and lastly by failing to deal with evidence (including his own) which might falsify his hypothesis.

The "innateness" of the P_s curve is essentially irrelevant, except as it underlies Lieberman's hypothesized "air pressure perturbation effect", and will therefore be discussed in connection with that sub-hypothesis. As for the matter of infant cries, the evidence here is badly obscured by several different varieties of misreporting by Lieberman. First, he refers to Bosma, Lind, and Truby (1964), which is essentially a report on cinefluorographic studies of infant pharyngeal movements during crying, and says virtually nothing about the question of "lungs versus larynx", i.e., about the relative importance of pulmonary and laryngeal activity in infant crying. The proper reference is Bosma, Truby, and Lind (1965), and even its contents are misreported. The study does not give direct evidence on as many cases as Lieberman's second-hand report would indicate (cf. Ohala, 1969), and Lieberman fails to report a significant finding of the authors (corroborated by Ringel and Kluppel, 1964) that "The expiratory volume changes as seen on the spirogram show a great variety in pattern, not only among different individuals .... but also within the same infant." (Bosma, Truby, and Lind, 1965: 73) Most serious is Lieberman's attempt to make it seem as though Bosma *et al.* were arguing the existence of "innate" and "archetypal" pulmonary activity: "The 'shape' of the fundamental frequency contours of the cries was similar to the shape of the typical esophageal pressure contour. Qualitatively speaking, the gross variations of the fundamental frequency contour thus seem to be a function of the subglottal air pressure during infant cries." (Lieberman, 1967: 43) In actuality, Bosma and his co-investigators were saying something quite different:

> The actions of the larynx and pharynx essentially define the infant's cry, since the less discriminate trunk motions of respiration are more or less predictable from the upper respiratory actions. (63)

Again, they comment that

> The laryngeal coordinations, manifested by accomplished sounds, are the most discriminate expression of this activity. The expiratory constrictions and inspiratory expansions of the pharynx and trunk are grosser expressions. (73)

And they add, in conclusion,

> In these perspectives, post-natal development of vocal expression may be described as the addition of upper pharyngeal and oral modulations to an already well-developed laryngeal vocal coordination. (89)

Thus, even if adult speech behavior is seen as a direct outgrowth of mechanisms involved in infant crying, this development would not argue in any way for the primacy of pulmonary behavior in the suprasegmental system, let alone for the "archetypality" of one type of intonational gesture as opposed to a different type. Furthermore, the connection between infant cry and adult speech seems a dubious one. It would seem more sensible to posit such a connection between adult behavior and the cooing and babbling vocalizations of infants, which are also manifested quite early. Lenneberg (1967) carefully distinguishes

> two distinct types of vocalization ... . The first type includes all sounds related to crying. It is present at birth (and potentially present even before the end of normal gestation). It undergoes modifications during childhood and then persists throughout life. These sounds as well as other sounds more immediately related to vegetative functions seem to be *quite divorced from the developmental history of the second type of vocalization, namely all of those sounds which eventually merge into the acoustic productions of speech.* (276; emphasis mine)

In differentiating between these two modes of vocal behavior, Lenneberg notes that "... cooing contrasts with crying in that it shows resonance modulation almost at once in addition to fundamental frequency modulation. In other words, during cooing some articulatory organs are moving (mostly tongue), whereas during crying they tend to be held relatively still." (277)

We would therefore argue that the mere fact that cry behavior occurs somewhat earlier than other infant vocalizations does not justify arguments for any "archetypal" nature of speech behavior, particularly since those very infant cries do not manifest the primacy of pulmonary activity which Lieberman has claimed for them.

Archetypal Adult Intonations

We come now to what we consider to be Lieberman's major claim, that in their normal speech behavior adults make use of his posited archetypal gestures, and do so with such a preponderance of activity as to justify the labelling of those gestures as indeed "archetypal", and the labelling of other behavior as "idiosyncratic 'personal' articulatory patterns." (Lieberman, 107) Unfortunately, Lieberman's evidence is not merely insufficient, but actually contrary to his hypothesis.

The essential evidence offered by Lieberman for his claims concerning adult intonational behavior consists of an experiment conducted by Mead, Proctor, and Bouhuys (1965), in which four male speakers of American English each recorded a list of sentences and words while seated in a sealed body plethysmograph, enabling the investigators to obtain records for the relative volume of air in the lungs, in addition to measurements of subglottal pressure (measured esophageally) and of fundamental frequency of phonation, which were then lined-up with quantized spectrograms, the lining-up process having an accuracy of ± 40 msec. Lieberman concludes that the experimental data show that "The tension of the laryngeal muscles for the unmarked American English breath-group appears to remain relatively steady throughout the sentence. The fundamental frequency of phonation is thus a function of the subglottal air pressure function, and it falls during the last 150-200 msec of phonation." (104) However, this conclusion is vitiated by several facts. First, there is Lieberman's own admission that

> The points in Figure 4.35, where fundamental frequency is plotted with respect to subglottal air pressure, have a fair amount of horizontal dispersion, which indicates that the laryngeal tension is not always constant throughout the non-terminal portion of each breath-group. Our initial hypothesis regarding the complete absence of variations in the tension of the laryngeal muscles during the production of a declarative sentence in American English must therefore be considered a first approximation. (102-3)

Despite this damaging admission, the summary on the very next page of his book, as we have seen, continues to maintain that f_o must be considered a function of P_s. Making this all possible is Lieberman's unique style of scientific argumentation, as seen in the following note:

> It is important to note that we are not claiming that the normal breath-group always has a uniform laryngeal tension. Its *archetypal* articulatory correlate is a uniform laryngeal tension that results in an acoustic output where f_o is a function of the subglottal air pressure. The speaker may use alternate articulatory gestures to produce an acoustic output that is similar

> to the acoustic output of the archetypal articulatory correlate. (96)

In Lieberman's defense, it must be said that he appears to believe that his "air pressure perturbation effect" hypothesis, which he uses in his interpretation of the Hadding-Koch and Studdert-Kennedy experiments, justifies this conception of a physiological world in which perception routines founded upon "archetypal" gestures can negate the importance of the physiological reality of the moment. However, in defense of the impartial reader, it must be said that if one rejects the "air pressure perturbation effect" hypothesis (cf. pp. 22-25 below), then we are left with an extremely ingenuous argument. But even this state of affairs might be partially acceptable if it were not for the serious deficiencies of the very data Lieberman presents for the verification of his hypothesis. Having entirely eliminated one of the four subjects from his discussion (thereby throwing out one-fourth of his data) because that speaker "produced exaggerated effects" (66), and having carefully selected appropriate examples for purposes of illustration (seven examples from Speaker 1, six from Speaker 2, eight from Speaker 3), Lieberman is still forced to ignore or explain away physiological facts, in order to defend his position that f_o is archetypally a function of P_s in unmarked breath groups. For example, in Figure 4.14 Lieberman's caption claims that Speaker 1 placed prominence on the word *Joe* by means of increased subglottal air pressure. However, the f_o peak for *Joe* comes c. 150 msec after the P_s curve peak. In those 150 msec, f_o *rises* c. 100 Hz, while the P_s curve *falls* by c. 2 cm H_2O. The true situation is obscured by the fact that the caption directly under the spectrogram is misaligned. When examined closely, it is clear that the peak amplitude on the diphthong of *Joe* correlates with the f_o curve peak, but not the P_s peak. It seems necessary to posit laryngeal action to explain this prominence peak.

In Figure 4.16, showing how Speaker 1 read the sentence "The number that you will hear is ten" twice on the same expiration, Lieberman measures the f_o and P_s curves at four arbitrary points, and thereby makes the f_o curve for utterance B seem lower than it really is in relationship to that of utterance A (thus suggesting that the second utterance, with its slightly lower P_s curve, must have a lower f_o), but it is still quite obvious that the speaker is utilizing a great deal of laryngeal tensioning. particularly in utterance B. Again, in reference to Figure 4.18, Lieberman admits (80) some laryngeal slackening at the beginning and end of the utterance ("Joe ate his soup"), but ignores the fact that the f_o rises in the middle of the utterance (on "ate") while the P_s curve is falling very sharply.

Before dealing with our final examples of faulty analysis by Lieberman, we must focus our attention upon his rather unsettling calculations for the ratio of f_o variation to variations in P_s. On page 71, Lieberman reports a ratio of 16-20 Hz per cm of water.

Elsewhere, he reports even higher values, and in his discussion of the Hadding-Koch and Studdert-Kennedy experiments (Lieberman, 1967: 99) he makes use of his figure of 20 Hz/cm H_2O in order to gain support for his particular interpretation of their results. However, such a figure is clearly beyond the pale, in light of many recent experiments performed precisely for the purpose of studying this ratio of f_o to P_s. Ohala, in his recent dissertation, averaged the results of 161 chest pushes, and obtained a figure of 2-3 Hz/cm H_2O for the pitch range used in speech. His comments on this general question of f_o/P_s ratios deserve repetition:

> These values compare favorably with those of Ladefoged (1963) and Öhman and Lindqvist (1966) who worked with living subjects, and are at least in the range of values reported by van den Berg and Tan (1959) and Anthony (1968) who worked with excised larynges, the differences being attributable to individual variation or to the different experimental conditions (living subject versus excised larynx). However, the difference between the values found in this study and the values derived by Lieberman (1967) from running speech, namely 18-22 Hz/cm. aq. are too large to be attributed to individual variation. This is not surprising, however, since Lieberman did not do any experiment of the kind reported above, and in fact made no attempt either to control or to monitor the highly relevant variable of laryngeal adjustment. He merely assumed that the laryngeal tension was constant. We have already shown that this is not a valid assumption. (Ohala, 1969: 78)

Further refutation is supplied by Vanderslice (1969), who says,

> Furthermore, Ladefoged (1962) and Öhman and Lindqvist (1966) have shown that f_o is a far weaker function of transglottal pressure drop than Lieberman would have us believe; he dismisses their results, speculating that "some sort of feedback control ... may function during singing (p. 97n)." But Öhman and Lindqvist used spoken, not sung, sentences, and Ladefoged anticipated and refuted the "feedback" objection. This factual discord would have warned an experimenter less infatuated with preconceptions to re-examine his logic. (Vanderslice, 1969: 3)

Lieberman's embarrassingly high f_o/P_s ratio is a heavy enough burden for his theory to bear. However, there are cases in which he necessarily closes his eyes to relationships between his f_o and P_s curves which would yield even more incredible ratios. Thus, in Figure 4.31 ("Did Joe eat his *soup*?"), Lieberman ignores a rise of c. 70 Hz in the f_o curve on the word "Joe", which takes place while the P_s curve remains flat. In Figure 4.32, Lieberman attributes the prominence on "*Joe*" in "Did *Joe* eat his soup?" to duration, ignoring a c. 35 Hz rise in f_o which occurs while the P_s curve is falling slightly. More serious is the case of Figure 4.27, which illustrates Speaker 3 reading the sentence "Joe ate his *soup*." Lieberman's total commentary for this figure reads, "Normal breath-group, prominence on the word *soup*." In his brevity, he overlooks some very troublesome data. During the enunciation of the word "*soup*", the f_0 drops 140 Hz in c. 45 msec, while the P_s drops by only c. 1 cm H_2O. This, of course, results in a ratio of 140 Hz/cm H_2O. Even if we assume a serious mismatch (with respect to time) of the curves, and measure later on the P_s curve (where the drop is somewhat steeper), we still get a ratio of c. 100 Hz/cm H_2O. The improbability of such a ratio would lead us to hypothesize the necessity of significant laryngeal activity even if we did not already have significant data attesting to its importance in influencing fundamental frequency contours (Ohala, 1969; Ohala and Hirano, 1967; Vanderslice, 1967).

The above example is important in another respect. It illustrates Lieberman's tendency to ignore phenomena of pitch drop (cf. also Lieberman, 1968: C-4-4). This blind spot becomes particularly crucial in respect to "scooped" accents, i.e., accents in which prominence is manifested by means of a drop away from a preceding higher pitch, or a low scoop up to the next higher pitch. Lieberman's model makes no provision for any such accents, a shocking deficiency in view of Bolinger's (1958) findings concerning the significance of such accents in American English. An example of this blind spot is Lieberman's analysis of Figure 4.24 ("Did *Joe* eat his soup?"). He is concerned with the duration of the stressed "*Joe*" (300 msec, as compared to 250 msec for another stressed "*Joe*", and lower values for unstressed examples). But this is clearly an artifact of the strong scooped accent, in which there is an exact match between a high peak in the P_s curve and a definite scoop in the f_0 curve. We mention this example because our data confirm those of Bolinger. In Chapters 4 and 5 we will show fairly frequent occurrences of such scooped accents, and we will also show that they were clearly understood by the listeners in our experiment.

Perception and "Analysis-by-Synthesis"

We mentioned earlier that in Lieberman's argument it did not matter whether a great preponderance of adult speakers actually used his hypothesized archetypal intonational gestures (but see pages 24-25 below), because he maintained that, no matter how intonations were produced, they were perceived according to routines incorporating those archetypal gestures. The evidence set forth in support of this extremely strong claim consists of Lieberman's reinterpretation of data published by Hadding-Koch and Studdert-Kennedy (1964). In their very stimulating experiment, the carrier phrase *For Jane* was processed through a Vocoder to yield forty-two different intonation contours. Each contour began at 250 Hz and remained level for 140 msec (making a neutral pre-contour of the word *For*). In the next 100 msec the contours rose to either 310 or 370 Hz. From this peak, the contours fell, over the next 200 msec, to a "turning point" of 130, 175, or 220 Hz. In the last 200 msec, the contours proceeded to one of seven "end points": 130, 145, 175, 220, 275, 310, or 370 Hz. Thus, each contour had a total duration of 640 msec. The subjects (24 American and 25 Swedish undergraduates) were presented with these stimuli in two separate sessions. In one, they were asked to characterize each contour semantically, i.e., to categorize it as a statement or as a question; in the other, they were asked for psychophysical judgments, i.e., whether the contour ended with a rising or falling pitch.

This experiment (only the first of a projected series) yielded a great deal of interesting data. In his analysis, Lieberman concentrates on one effect. It happened that American listeners gave identical psychoacoustic and semantic responses (80% "rising", 80% "question") to two somewhat different contours. One rose from the 250 Hz precontour to a peak of 310 Hz, fell to a turning point of 175 Hz, then rose to an end point of 335 Hz. The other contour had a peak of 370 Hz, a turning point of 175 Hz, and an end point of 265 Hz. Lieberman asked why a much smaller terminal rise (in the case of the second contour) was able to elicit the same judgment from the listeners, and supplied an answer based upon the "analysis-by-synthesis" version of the "motor theory" of perception. This theory, which states that speech signals may be perceived, not according to acoustic phenomena, but according to the listener's estimate of the articulatory efforts which produced those acoustic effects (thus, analysis-by-synthesis), has been advanced by various investigators (Liberman *et al.*, 1963; Halle and Stevens, 1964), partially supported by others (Ladefoged, 1962a; Ladefoged and McKinney, 1963; Lieberman, 1963; Kozhevnikov and Chistovich, 1965; Galunov and Chistovich, 1965), and attacked by others (e.g., Lane, 1965, 1967). In Lieberman's version, the high "question" rating for the contour with the slight terminal rise results from the listeners' estimating the additional subglottal pressure which must have been expended in producing the higher peak (370 Hz), knowing through experience that such pressure increases early in a sentence result in a generally lower pressure later, and thereby expecting the same degree of laryngeal tensioning (at the end of this "marked" breath-group) to produce a lesser rise. He calls this the "air pressure perturbation effect."

There are various objections that can be made to Lieberman's interpretation of the so-called "peak effect" in the Hadding-Koch and Studdert-Kennedy data. Ohala objects by reasoning that

> ... a given contour would be more likely to be identified as a "question" if (1) it had a large rising pitch at the end, or if (2) it had a large rising pitch *before* the end and (a) thereafter remained high and possibly (b) had a slight rise at the end. This is exactly what the results of the Hadding-Koch and Studdert-Kennedy data reveal. When the pitch rises to the 370 Hz high point this is indeed interpreted by the listeners as "prominence" on that syl-

> lable -- Lieberman is right on this point. However, what Lieberman does not recognize is that a pitch rise early in the utterance due to prominence or emphasis manifests the major intonation contour of the sentence ... . This accounts for the fact that a smaller terminal rise is sufficient for the 370 Hz high point stimuli to be judged as questions. This first large pitch rise is in itself a strong cue for "question" and seems to tip the perceptual scales in favor of that judgment; another large pitch rise at the end is not necessary for the contour to remain "question"-like. In order for a stimulus with such a large pitch rise to get a "statement" judgment it is necessary that there be an extra low pitch fall immediately thereafter and that the pitch remain relatively low.
> (Ohala, 1969: 122-23)

Even more damaging refutation comes from the orginal investigators themselves. In a report on their later experimental work, Studdert-Kennedy and Hadding-Koch suggest that much of the importance of the "peak effect" might have been due to the nature of the stimuli in the first experiment:

> In our earlier study (Hadding-Koch and Studdert-Kennedy, 1964, 1965), the peak effect was clear in both semantic and psychophysical judgments of the two language groups, though the Swedish were less consistent in their psychophysical judgments than the North Americans. Here, the effect has disappeared almost completely from the psychophysical data of the Swedish group and is only marginally present for the American ... this suggests that the effect is linguistic rather than psychophysical, and the problem is then to explain why the effect appeared in the psychophysical data of the earlier study.
>
> The factor most likely to be implicated seems to be the stimuli themselves. In the present study, with its utterance /nɔˈvɛmbə:/, the glide from precontour to peak to turning point lay on the second syllable, the terminal glide over the third syllable; in the earlier study the utterance /fæ ˈJeɪn/ was used, and the entire frequency sweep from precontour to end point lay over the single syllable /ˈJəɪn/. Listeners to the latter may have found it difficult to separate perceptually the glide to the peak from the terminal glide, and so were inclined to assign a higher value to the terminal glide when the peak was high than when the peak was low. (Studdert-Kennedy and Hadding-Koch, 1969 forthcoming)

This interpretation seems reasonable, and is partially corroborated by Elekfi's experimental study of Hungarian intonation (see Juhasz, 1963), which found that when within a stress group the pitch changed direction, some listeners noticed only one direction (Elekfi's subjects were all trained phoneticians).

We believe that this later interpretation of their own work by Studdert-Kennedy and Hadding-Koch essentially disposes of Lieberman's elaborate theoretical construct based upon an isolated instance from their earlier experiment. However, there remain some basic questions relating to Lieberman's version of motor theory, and the relationship of this theory to other parts of his theoretical apparatus.

First, Lieberman's application of motor theory to the perception of fundamental frequency would seem to imply that f_o curves either constitute in themselves, or are composed of certain fixed categories. While the early work on the motor theory of perception by the Haskins group related perception to articulation of segmental phonemes, particularly consonant phonemes, there is no reason to believe that f_o curves are constrained within the same kind of range limitations as are consonant phonemes; furthermore, Lieberman's own 1965 critique of the Trager-Smith system would argue his own disinclination toward regarding intonations as made up of clearly sub-categorizable entities of pitch, stress, etc. Therefore, the only entity which might be thought of as "fixed", and therefore perhaps capable of functioning as a basis for an "analysis-by-synthesis" routine is the unmarked or "normal" breath-group itself, since Lieberman considers it to be "archetypally" related to the "innate" P_s curve, which is characterized by a rise to a peak, and by an abrupt fall during the last 150-200 msec. But the evidence here is quite shaky. We have already pointed out the unreliability of Lieberman's reportage of the data on infant cries, data which seems, upon closer examination, to argue *against* his conception of an innate f_o curve directly related to a P_s curve. A similar problem seems to attend his use of Armstrong and Ward (1926) and Cowan (1936) to demonstrate a strong tendency toward the use of unmarked breath-groups by adult speakers.

Lieberman cites cases from Armstrong and Ward to show a tendency for the pitch to fall "when the choice is at all possible." (169) But these cases, typified by the following examples,

He strolled aimlessly about the road ↘ kicking stones out of his path. ↘

She shook hands ↘ and said she was glad he had come. ↘

seem to be instances of *part*-falls (we are referring, of course, to the first fall in each sentence), and *not* of the full-falls necessary for Lieberman's unmarked breath-group. Such part-falls would be included within Hultzén's "not-low" category, and, within the extension of Hultzén's system which we shall formalize in Chapter 6, would be considered basically as variants of sustained contours.

In a footnote to his brief discussion of Cowan (1936), Lieberman makes the following comment:

> It is interesting to note that Cowan's instrumental study shows far more cases of nonterminal normal breath-groups than does the Armstrong-Ward perceptual analysis. Instruments cannot infer the presence of a cue from the structure of the language. Linguists, however, may "hear" acoustic signals where none exist. (170)

The second part of this quotation is quite true, and it is useful for linguists to be reminded of the possibility that they are supplying syntactic or other information to a speech signal. However, the first part (referring to "nonterminal normal breath-groups") appears to be based upon an unfortunate misinterpretation of Cowan's findings. Cowan found that "63 percent of all phrases ended with a falling inflection, 12 percent with a rising inflection and 25 percent with a level intonation." (Cowan, 1936: 81) However, it is important to note that Cowan is speaking of "falling inflections", and not of a category such as [full] fall. An inspection of his Figure 3 (75) and Table II (76) makes it clear that most of these falling inflections would have to be placed in a "part fall" category, and that quite a number of them would probably not be perceived as falls by listeners in an experiment of the type performed by Hadding-Koch and Studdert-Kennedy. It is therefore quite inaccurate for Lieberman to refer to all of these inflections as examples of "nonterminal normal breath-groups."

There are other questions which might be raised concerning the relationship between utilization of vital capacity (the breathing capacity of the lungs) and the kind of internal computation envisioned in Lieberman's analysis-by-synthesis model. For example, would the analysis-by-synthesis of an intonation contour heard by an Olympic long-distance runner be different from that heard by a ninety-seven pound weakling? Lacking any data on this matter, we are in no position to accord to Lieberman's analysis-by-synthesis model of intonation perception the same degree of confidence merited by the findings of the Haskins group (cf. Liberman, *et al.*, 1963) concerning the perception of consonants.

A Tentative Theory of Intonation

In the preceding section we found that the available evidence does not support the kind of simple, physiologically-conditioned model for intonation put forward by Lieberman. We therefore suspect that those quasi-universal elements of intonation observed by Bolinger and others will have to find their interpretation through precisely those vague, mentalistic notions (e.g., "finality" versus "continuation") which we earlier suggested would be extremely difficult to prove in any definitive fashion. We leave that task to others, though perhaps some of our data on American English intonation patterns may prove to be of interest to those working on universal aspects of intonation. In our own study of the intonation of American English, we have found especially instructive the work of Bolinger and of Hultzén. Bolinger's contributions cannot be said to constitute a total theory of intonation, but his many insightful articles have given us a clearer idea of the nature of English intonation, and his major work (Bolinger, 1958) has shown the power of careful experimentation in elucidating intonational phenomena. Some of his results have been mentioned earlier in this chapter.

Hultzén's work, while lacking significant experimental validation, proposes a model flexible enough to deal with many important aspects of intonation, and deserves to be rescued from the neglect into which it has fallen. Hultzén begins with a principal distinction between "accented syllables, *accents*, and weak accented or unaccented syllables, *unaccents*." He adds that

> It is generally true that accents have louder stress than unaccents, but the loudness of accents of the same degree is not by any means uniform. In general the accents, or the vowels in accented syllables, are longer than unaccents, again not uniformly. Accents do not regularly have higher pitches than unaccents; accents and unaccents have regular pitch places in the intonation pattern. All of this could as well be described in terms of stress and its influence on length and pitch; but to do so would be to assume a primacy for stress which I do not think can be demonstrated. (Hultzén, 1957: 319)

Hultzén divides intonation into what he calls three different patterns, which we might think of as different strata, since one seems to build upon another. The first pattern he calls "formal". As he defines it, "The formal pattern is the interior arrangement of the accents and unaccents." (Hultzén, 1957: 320) The formal pattern communicates only the fact that the speaker is a native speaker of, for example, General American (For partial corroboration of the communicative significance of this stratum, see Atkinson, 1968). The next stratum is the syntactic pattern. Hultzén comments,

> The syntactic pattern is primarily in the shape of intonation at phrase end. The communication is the place of the phrase in the larger structure [the sentence] ... A not-low phrase end correlates with noncompletion and a low phrase end with completion. (320)

The superimposition of the syntactic pattern upon the formal pattern gives us the carrier tunes, one "open" (incomplete), as the other "closed" (complete) (see Figure 2.1). He calls these tunes "colorless carriers of acceptable speech".

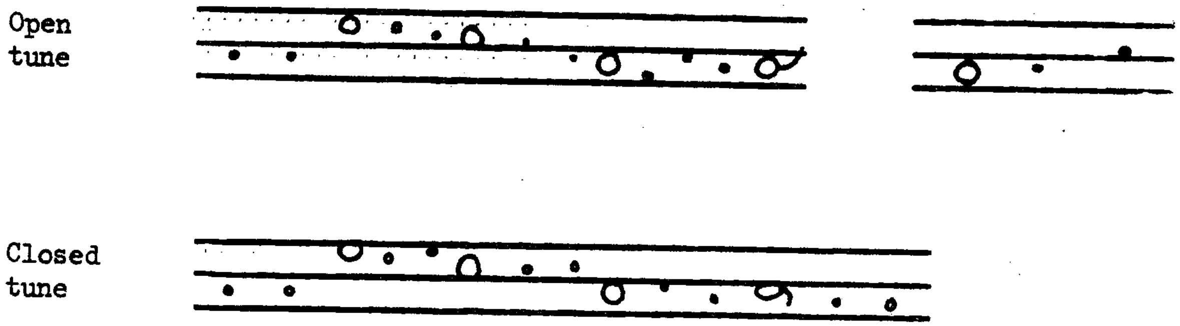


Figure 2.1: Hultzén's "carrier tunes". (After Hultzén, 1957: 317) Note the resemblances to the two-tune British analyses of Armstrong and Ward (1926) and others.

Hultzén describes the third stratum, the rhetorical pattern (essentially emphasis) as follows: "The signal in the rhetorical pattern ... is some modification of the carrier tune ... . Any modification will do, but by far the most usual is a strong accent, with higher pitch and greater stress and greater length, with or without a downturning on the accented syllable of the significant lexical item." (326-27) He summarizes his system in the following manner:

> The essence of the theory rather sketchily presented here is that any phrasal intonation is to be interpreted, not as a single pattern having in it a specifiable communication, but as three patterns which are superimposed one upon another and each of which has in it a specifiable communication. For the most part the communication in any one pattern is fairly simple. For the formal pattern it is that the speaker is speaking the dialect as it is spoken or is not. For the syntactic pattern it is that the sentence is complete or is not complete. For the rhetorical pattern it is that what is being said at any point is important or is not important, with some suggestion of the degree of importance. (328-29)

As it stands here, this model is capable of distinguishing between native and non-native speaker performance with respect to the accent system of the language, between finitive and continuative clauses, and between emphatic and normal utterances or portions of utterances. It thus covers more territory than at first appears to be the case. It does not propose a systematic coverage of the "emotive" aspects of intonation, but such coverage is rare in any attempt at a systematic description of English intonation (In Chapter 5 below we show how paralinguistic cues interact with variations in contour shape to produce effective emotive intonational gestures in English). However, its major deficiency would appear to lie in the area of syntactic communication.

One might surely object that the distinction between finitive and continuative is too gross to suggest the possible number of contrasts, and that a more finely graded system is necessary. We share this belief, but in Hultzén's defense it must be said that his refusal to enlarge the number of contrasts was a principled one:

> Although I show the shape of the syntactical open tune as rising, I always speak of the end as not-low. The reason is that I do not believe, as some do, that there is a contrast at the level of generalization represented by these GA carrier tunes between any two of the patterns ... [rising, sustained, or falling]. Since some of the current theories of intonation do make a distinction between end-of-phrase upturn and end-of-phrase sustained or level, a difference which is certainly perceptible, I feel it necessary to explain that I have made considerable effort to find some corresponding difference in the communication and have not been able to do so. It is one thing to observe a difference between A and B which might constitute a significant contrast, and quite another thing to find that all instances of A occur in connection with a communication, syntactic of other, which is actually in contrast with a communication in connection with which all instances of B occur. Some speakers seem to have A and B, or even A and B and C [i.e., rising, sustained, and falling] in free variation. (325)

The above passage may be considered the starting point of the present investigation. Hultzén suggests that he attempted controlled experimentation, and that he was unable to find consistent communicative contrasts. But suppose that one were to begin with a somewhat different hypothesis. Suppose that one assumed that although many speakers might not make a consistent contrast between two intonational contours (e.g., end-of-phrase rising versus end-of-phrase sustained), *some* speakers do make such a consistent contrast, and make it clearly enough so that a large group of listeners, *including speakers who do not make such a contrast,* can correctly categorize those contrasted intonations. We made this assumption, and a further one as well: we assumed that the relevant variable would be socio-educational (i.e., consisting not merely of courses of study, but also of leisure-time activities associated with those areas of study; in the case of graduate students in speech, this might involve work in oral-interpretation performance groups; in the case of graduate students in English literature, this might involve attending or listening to performances of classic works of English and American drama). Using a carefully-selected group of speakers, as well as a control group of speakers lacking the special educational background of the first group, we found significant differences in intonational performance in respect to both the number of intonational contrasts made, and the relative success of the speakers in communicating particular contrasts.

A report on our experimental procedures is given in the next chapter. Chapter Four contains data on syntactic elements in American English intonation. Data on emotive aspects of intonation will be found in Chapter Five.

Chapter 3: The Design of an Experiment for Studying Intonational Contrasts

Experimental Procedures

Our basic approach follows attitudes suggested by Bolinger and Hadding-Koch. Bolinger has said,

> For linguists who wish to be scientists there is only one scientific procedure: *gather the facts first,* morphemes, or profiles, that speakers are heard to use, together with the implications (determined by responses and otherwise) that these carry. To take as the starting point anything but such facts is to write with the mind open, perhaps, but with the ears closed. (Bolinger, 1949: 253)

Along similar lines, Hadding-Koch has suggested that

> Another way [of studying intonation] might be, in cases of unanimous responses on contours, to study the corresponding curves obtained by means of instruments, the sonograph, mingograph or others, in order to see whether any corresponding agreement can be found there. (Hadding-Koch, 1956: 90)

We thus chose to perform an experiment in which examples of American English intonation contours would be presented to listeners, who would be asked to categorize them on a forced-choice basis. There remained, however, several basic questions of procedure:

(1) What kind of speech? The essential choice was between natural and synthesized speech. Although it is clear that synthetic speech can be far "neater" to work with (cf. Hadding-Koch, 1961 versus Hadding-Koch and Studdert-Kennedy, 1964), there is nonetheless a certain epistemological sleight-of-hand involved in extensive experimentation with synthetic speech, because in planning his sample of synthetic speech, the experimenter is obviously building upon a model of the natural language system, whether that model is explicit or implicit (it being unlikely that an experimenter would present his subjects with four cycles of a sinusoidally varying frequency as a sample of typical English intonation). Since this is so, most investigators would agree that we should not continue to test samples of synthetic intonation

patterns without at least occasional glances back to the natural language which we are supposedly replicating. An example of the dangers inherent in a divorce between natural and synthetic study of intonation can be seen in the work of Isačenko and Schädlich (1963), in which extremely simplified synthetic contours were presented to listeners. We agree with Hadding-Koch's criticism of their work that "... one cannot be sure that the contours chosen by the listeners represent relevant features of natural speech -- only that these particular items were chosen by listeners among those presented, in that particular situation." (Hadding-Koch, 1964: 130) We therefore concluded that there was value in attempting controlled experimentation with carefully elicited samples of natural speech.

(2) What kind of samples of natural speech? The essential choice here lay between sentences of normal length and short words or phrases which were, in most cases, capable of standing as sentences. Among others, Hadding-Koch has remarked that "it is, however, very difficult to find listeners who react rapidly and adequately enough to have any opinion whatsoever on the fast speech of ordinary conversation." (1956: 90) Some observations of our own support this conclusion. In an earlier experiment attempting to assess the fit of Halliday's contour system with American English intonation, we found that listeners had great difficulty in dealing with an intonation contour spread over a sentence of moderate length. There are at least three ways of attempting to diminish this difficulty. One might use only trained phoneticians as subjects (cf. Juhacz, 1963), but with no guarantee that they would constitute a valid, or even especially competent (cf. Ladefoged, 1962b) sample of the population. One might instruct listeners to focus their attention on a particular portion of the utterance, but to do so is really to judge the matter in advance of the experiment. Or one might follow Uldall (1964), who allowed listeners to have sample utterances repeated as many as fifteen times. Such a procedure, however, can hardly be said to duplicate normal language hearing conditions.

There is an additional problem attending any attempt to present moderately long sentences with varying intonation contours to listener-subjects. We refer to those listeners' expectations based upon non-intonational elements such as word order. Thus, although Uldall remarks (1962: 799) that "It is of course a commonplace that in many languages, including English, sentences need not be cast in a special question form to operate as questions," there is nonetheless some "loading" of expectations here, deriving from the use of statement word order with a particular string of words. This loading, which would in this case operate *against* a "question" judgment, can come from at least two sources: (1) Only the kind of questions which Bolinger (1957) calls "repetitive" function with statement word order, and they are by no means the most frequently used type of question in English. (2) The paraphrase possibilities for a "question" interpretation are far greater in this case than for the statement. Thus, "He'll be here on Friday" occupies a large central portion of the semantic-paraphrase "space" for this proposition (as compared with, for example, "It's on Friday that he'll be here"). On the other hand, "He'll be here on Friday?" is far from the center of the semantic-paraphrase space for the questioning of X's possible arrival on Friday. Far more central to this space would be locutions such as "Do you expect him on Friday?" or "Is he coming on Friday?" or "Is he going to be here on Friday?" or even "Did you say he'll be here on Friday?" Of course, one function of the repetitive question is to imply the construction "Did you say *S*?" However, in Uldall's experiment, the intonation is being forced to do *all* the work, and we have no empirical basis for assuming this to be the normal, "expected" situation.

For these reasons, it seemed far preferable to avoid using moderately long sentences in the samples of speech presented to our listeners, and to attempt instead to choose brief speech segments which would nonetheless be capable of bearing many kinds of intonation contours with little or no bias. As a result of some preliminary experiments, we chose three kinds of segments: the word *yes*, the word *ridiculous*, and the phrase *were they black*. The great majority of the speech samples elicited consisted of *yes*, uttered in various circumstances. In choosing this word as a neutral and flexible "carrier phrase", we were in effect following the example of Daniel Jones, who illustrated the following six varieties of meaningful intonation with the word *yes*:

(1) low fall	meaning "that is so"
(2) high fall	meaning "of course it is so"
(3) rise-fall	meaning "most certainly"
(4) (somewhat) high rise	meaning "is it really so?"
(5) low rise	meaning "yes, I understand what you have said; please continue" (the telephone *yes*)
(6) (rise) fall-rise	meaning "it may be so." (1956: 151)

This pattern can be compared to one noted by Isamu Abe (1957-58: 183) in Uldall's unpublished 1939 dissertation for intonation patterns accompaning the nasal sound /m/, e.g. Rising (= Yes. Go on.); Falling-rising (= Yes, doubtfully); Rising-falling-rising (= Yes, but ...); Level-rising-falling (= How impressive!).

The Nature of the Dialogue

In designing our experiment, we constructed a dialogue, in which were "buried" several test samples of the word *yes*:

Yes[1]	placed in an environment designed to elicit a simple statement intonation contour.
Yes[2]	placed in a "Yes, but ..." environment. Expected to show the intonation of an incomplete statement.
Yes[3]	Jones' "please continue," or "telephone" *yes*.
Yes[4]	Jones' "is it really so?"; Bolinger's "repetitive" question, meaning "did you really say *yes*?"

These four samples were later matched against four similarly-conceived *yes*'s elicited by means of cue cards.

The dialogue also included presumable unemphatic and emphatic utterances of the word *ridiculous*, also later matched against similar samples obtained by means of cue cards, and three instances of the phrase *were they black*. The reason for the inclusion of this latter item again derives from the writings of Hultzén. As we noted on page 28, Hultzén expressed considerable doubt that speakers of General American made a consistent contrast between such not-low phrase end contours as rising versus sustained. In another of his papers, he extended this notion:

> Other formally marked non-finitive texts are initial clauses ... introduced by *after*, *if*, *when*, etc. ... And certain word orders ... [e.g.]: (1) "Were they better, they'd be more acceptable." (2) "Were they better, or worse, than you expected?" (3) "Were they better?" Although in (3) the clause is printed with end pointing, the text shape within the clause, the matter in hand, is the same as it is in the other contexts, where obviously non-finitive, *and the basic intonation is the same in all three*. It is at least surely so that the open intonation has been established in English for this clause shape, occurring very frequently in the (3) setting. In some idiolects this non-finitive text shape has an arrested down-turn or slight up-turn in situations (1) and (2) but an extensive up-turn in situation (3). *These forms can be considered positional variants rather than two different intonations*." (1964: 87 Emphasis mine.)

Bolinger apparently held the same view, for in his "Theory of Pitch Accent in English," he had shown exactly the same (rising) contour on the *were they better* sections of the two utterances "Were they better?" and "Were they better they'd be more acceptable." (1958: 147)

It is a well-known fact that the very ease and fluency of adult language behavior poses problems for those attempting laboratory experiments upon such behavior. When the task is too easy, it becomes next to impossible to obtain rankings of proficiency among those performing the task. For this reason, much contemporary study of speech behavior incorporates artificial handicapping devices such as delayed feedback or masking noise (cf. Miller and Isard, 1963). Because the listener's (and dialogue-reader's) expectations upon first encountering a string such as *were they ADJ* would be heavily biased in favor of a simple question interpretation, as opposed to a possible *or*-question (*were they ADJ or ADJ?)* or a subjunctive interpretation (*were they ADJ* [*then*] *S*), and because of the great rarity of the subjunctive usage in typical spoken English, it was felt that this possible three-way contrast would constitute a good test for separating the "sophisticated" speakers (i.e., those with a rich system of actually-utilized intonation contrasts) from the "naive" or less-sophisticated speakers. This was considered to follow from the suppositions that, first, the speaker would have to be aware of the different patterns, and secondly, the speaker would have to make the contrasts clearly, in order to overcome the listeners' bias in favor of the simple question interpretation of the string. It turned out that the contrast was more common than we had believed, and far more common than Bolinger and Hultzén had maintained; however, this was one case in which the skill in communicating the contrast did not match with the hypothesized "sophisticated"/"naive" split. (See pages 58-68 below for further discussion of these contrasts.) One change was made from the Bolinger-Hultzén example. In a pilot study, it was found that the ending on *better* tended to run together with following sounds, particularly in a phrase such as *better or worse*, so the sample phrase was changed to *were they black*, in order to ease the problem of editing, which was already a serious one.

The dialogue read by the speakers is reproduced on the following pages. *Yes*[1] is the first *yes* (line 2). *Yes*[2] is the *yes* in "Yes, but ..." (1.9). *Yes*[3] is in 1.18. *Yes*[4] is in 1.37 (thirteen lines from the end of the dialogue). The reader will notice that "stage directions" were supplied for *Yes*[3] and *Yes*[4]. It was felt that the danger of biasing the speakers was less serious than the danger of exposing the listeners to speech samples produced by speakers who did not understand the dialogue. Aside from these two stage directions and the general directions printed at the beginning of the dialogue, the speakers were given no advice on how to read it, and were given only a few minutes of silent perusal before recording the dialogue. Below is the full text of the dialogue, as it was read by the speakers:

Preliminary directions: Please say your name and the date. Then read the following sentences into the microphone:

1. "This is speaker number ______."

2. "Today is Monday."

Directions for reading the dialogue: You will take both "parts" in reading this dialogue. Do *not* try to use any difference in voice between "A" and "B". Read them both in your natural voice, and do *not* try to over-act. Read the parts as though you were taking part in a natural, somewhat spirited conversation.

A. Hi -- You're Jim's friend, aren't you?

B. Yes.

A. I think I met you at his party last week. We both got into that discussion they were having. It was a little bit ridiculous.

B. Don't worry about it. I thought you were one of the more sensible people there.

A. Thanks. Did you see Jim today?

B. Yes, but we didn't have much time to talk.

A. Oh -- That's too bad.

B. Why? What's the matter?

A. This is going to sound a little silly to you, but I always ask him to interpret my dreams, and I had a wild one last night.

B. What was it like?

A. Well, to begin with ... there were these *cats* walking around ...

B. (urging "A" to continue) ... Yes?

A. Well, it's just that I was bothered by the thought of all those cats.

B. Were they black?

A. What do you mean?

B. I mean, were they black or were they some other color?

A. What difference would that make?

B. Well ... were they black, they'd symbolize bad luck.

A. Why, that's ridiculous! How can the color of a cat mean so much?

B. I thought you believed in dreams!

A. Oh, I guess I do, but within limits.

B. Frankly, I don't know much about them, either. You'll probably have to talk to Jim.

A. I suppose so. Where are you off to?

B. I'm going to try and see Professor Anderson.

A. What about?

B. Just to talk. You see, Stanford made him an offer, and he said yes.

A. (not quite sure what "B" has just said) ... "Yes"?

B. That's right. He agreed to go there. I guess his department was giving him a hard time, so the Stanford offer looked good.

A. It seems as though every time we get somebody good, we lose him.

B. I know. Let's not talk about that, or we'll just get depressed.

A. I guess you're right. Anyway, it's been good talking with you. If you see Jim, let him know I'm looking for him.

B. Okay, I'll do that. So long, now.

A. So long.

There was one major change in the nature of the dialogue from our pilot test to the present experiment. In the earlier version, the experimenter joined the speaker in the recording booth and read part "A" (of a somewhat different dialogue) while the subject read part "B". However, following a suggestion of Professor Ilse Lehiste, we adopted the procedure used in Hadding-Koch (1961), in which the subject read both parts of the dialogue, while the experimenter confined his activity to monitoring the recording outside the recording booth. This modification in procedure produced much more natural readings, with less danger of "contamination" by the reading style of the experimenter.

The Nature of the Cue Cards

The procedure of elicitation by cue cards was intended to serve two purposes. First, as a check on test items elicited by means of the dialogue. Second, as the simplest method of eliciting samples in which some degree of emoting (e.g. "calm" versus "angry" readings of the word *yes*) was required. Regarding the first, we had double samples (from both dialogue and cue card elicitation) of *Yes*^{1}, *Yes*^{2}, *Yes*^{3}, *Yes*^{4} and the emphatic and unemphatic readings of *ridiculous*. The double samples of *ridiculous* correlated well, and we chose to use the cue card versions, since they were "cleaner" acoustically. There was, however, a problem with the matching of the dialogue and cue card samples of *Yes*^{3} and *Yes*^{4}. Apparently the instructions were confusing, so that some subjects who had uttered very natural intonation patterns for *Yes*^{3} and *Yes*^{4} during their dialogue readings nonetheless produced quite anomalous patterns from the cue card instructions. Although no such problem existed in the case of the cue card elicitations of *Yes*^{1} and *Yes*^{2}, it was decided to discard the cue card samples for *Yes*^{1} through *Yes*^{4}, and to use only the dialogue elicitations of those items for the listening tests.

Only cue card elicitation was used for the remainder of the test items. Judging from the results of the listening tests, this procedure worked quite well. Items theoretically commanded by the entire population (e.g., emphatic/unemphatic) were performed well by all subjects. On the other hand, items which required ability to emote vocally were performed well only by those who possessed that ability.

The texts of the cue cards are reproduced below:

Cue Card #1 (*Yes*^{1})

> Please say the word "yes" as though you were answering the question: "Was the correct answer 'yes' or 'no'?"

Cue Card #2 (*Yes*^{2})

> Please say the word "yes" as though you were answering a question with the words "Yes, but ..." Do not pronounce "but". Just pronounce "yes" as though the next word following would be "but".

Cue Card #3 (*Yes*^{3})

> Please say the word "yes" as though you had heard a friend say something to you, and you wanted him to continue with his statement. In other words, say "yes" as though you meant it to mean "I am listening to you ... please continue speaking."

Cue Card #4 (*Yes*[4])

Please say the word "yes" as though you were *questioning* what your friend had just said, and wanted him to repeat it or otherwise confirm it.

Cue Card #5

Please say the word "ridiculous"

a) in a neutral, *un*emphatic manner

b) in a very *emphatic* manner

Cue Card #6

Please say the word "yes"

a) in a quite *un*emphatic manner

b) in a very *emphatic* manner

Cue Card #7

Please say the word "yes" as though you were:

a) very bored

b) very interested

Cue Card #8

Please say the word "yes" as though you were:

a) full of belief

b) full of disbelief

Cue Card #9

Please say the word "yes" as though you were:

a) very calm

b) angry

c) angry, but trying to contain your anger

Cue Card #10

Please say the word "yes" as though you were:

a) completely unafraid

b) very afraid

Note: The above item was discarded, partly to keep the listening tests from running too long, partly because most of the readings sounded quite artificial.

Cue Card #11

Please say the word "yes" as though your mood were:

a) very agreeable

b) very disagreeable

Cue Card #12

Please say the word "yes" as though you were:

a) a virgin

b) a man (or woman) of the world

Cue Card #12 was originally put in as an end-of-session joke, but retained when it started to produce very interesting data. However, because of its superficially unserious nature, and because the listening test for this item was quite different from that used with the other items, it will be reported outside this dissertation, in an article tentatively entitled "The Vocal Quality of Innocence and Worldliness."

The Nature of the Recording Sessions

As noted above, we intended to elicit from the speakers a number of intonational test items, by means of dialogue reading and cue cards. In order to keep the speakers from deducing the test items buried in the dialogue, it was necessary for all speakers to read the dialogue first, then proceed to the cue cards. A serious question of procedure was involved here as to whether it would be better to record each speaker in one session or two separate sessions, one for the dialogue reading, one for the cue cards. Like many problems in psychophysical testing procedure, this was of a Scylla and Charybdis nature. On the one hand there was the possibility that the subject would remember test items from the dialogue reading (even though those items had been made as unobtrusive as possible), and would thus be influenced in the reading of the items on the cue cards. On the other hand there was the well-known fact that the fundamental frequency range of a normal speaker's voice can vary not only from day to day, but from morning to afternoon of the same day, so that attempts to match dialogue elicitations with those from cue cards would be seriously hampered.

Of the two dangers, the latter, physiological constraint of fundamental frequency range variation was considered the more serious, and it was decided to record each speaker in one session, while resorting to the following anti-"learning" device: after the subject had read the dialogue, monitored the reading for "naturalness", and either approved the tape or re-recorded the dialogue in a manner more satisfactory to his own ears, he (or she) would be invited to take a break and chat with the experimenter for a few minutes. The experimenter would ask the subject questions about his academic major, future plans, and similar topics designed to keep the subject talking, and to steer his attention away from the dialogue which had just been read. Only after several minutes of this distraction would the subject be invited back into the recording booth to read from the cue cards. Although the subject was given an opportunity to peruse the dialogue for a few minutes before recording it, this was not the case with the cue cards. The cards were in an ordered pile, and the subject could not see any card but the one from which he was reading at the moment. After recording from all the cards, the speaker again monitored his reading to determine whether each utterance sounded "natural". If he rejected, say, two items, he was given only the cards for those items, and invited to re-record them to his own satisfaction. After the conclusion of the recording session, several weeks were allowed to pass before the speaker took the listening test, in which his utterances were mixed with those of the other speakers. In this way, it was hoped to obtain consistent speaking performance, and speaking and listening performance uncontaminated by learning effects.

The Nature of the Speakers

Twelve speakers were chosen for this experiment. Within each category or sub-category, half were male, half female. The entire group was subdivided into six "naive" and six "sophisticated" speakers, but this was only a rough approximation of how they were expected to perform. The hypothesized socio-educational background advantage was expected to apply only to those whose work in advanced courses in speech or English literature had involved them in performing or listening to performances of poetry, drama and artistic prose, and whose exposure to these performances had enriched their intonational systems by suggesting to them better ways of communicating syntactic or emotional ideas through intonation. Therefore, the hypothesized "sophisticated" group really consisted only of the two graduate students in English (Speakers 9 and 10) and the two graduate students in speech (Speakers 11 and 12). (Note that male speakers are designated by odd numbers, females by even.) In order to suggest that it was the study by the above subjects of speech performance and of literature which produced the expected advantage in intonational communication, two graduate students in linguistics (Speakers 7 and 8) were included in the "sophisticated" group, to round it out to six. These two speakers were not expected to do as well as the graduate students in speech and English. In other words, if the graduate students in linguistics did not perform as well as the other graduate students, then the results would suggest that the increased richness of intonational systems did not result from scientific study of the phonetics and phonology of various languages, but rather from the study within the English language of literature and ways of orally communicating its subtleties.

Of greater importance to our study than the question of who were the best communicators were two more general hypotheses: (1) that a considerable amount of syntactic and emotional information could be communicated by intonation alone, and (2) that some native speakers of English would be more effective than others at communicating this information. Data for our positive findings on these hypotheses is given below, in Chapters 4 and 5.

The naive group was deliberately balanced for age with the graduate students in the "sophisticated" group, and consisted of mature undergraduate students at UCLA, where the others were doing their graduate work. None of them had any experience with advanced speech or literature courses or with acting, or with advanced work in any foreign language.

With both groups care was taken to ascertain that they and their parents were native speakers of English and that no foreign language was spoken in their homes. The only slight exception was Speaker 11, whose father had been born in Norway, but came to the United States at the age of two. It is possible that Norwegian intonation patterns from two generations previous might have influenced Speaker 11 and accounted for an occasional anomalous reading, but this would be difficult to prove.

All of the speakers had lived all of their lives in portions of the country described in elementary textbooks as places where a "General American" dialect is spoken. No one who had lived in New York, Eastern New England, or the South was chosen as a speaker. Most had grown up in the Middle West and migrated to California with their families.

The list of speakers is given below:

Table I

Speaker #	Initials	Sex	Age
1	RS	M	26
2	MR	F	24
3	RB	M	24
4	DH	F	31
5	MA	M	23
6	IF	F	23

Average age of "naive" group: 25.2; median age: 24.

7	KT	M	25
8	MH	F	25
9	GS	M	25
10	BS	F	22
11	JC	M	24
12	JP	F	26

Average age of "sophisticated group: 24.5; median age: 25.

The Nature of the Listening Test

The twelve speakers were also, some weeks later, utilized as listeners. There were thirty-eight additional listeners (making a total of fity) with an average age of 21.20. Because many of the non-speaker listener subjects came from the subject pool of the Psychology Department, it was not possible to engage in pre-screening. However, at the beginning of the test they answered questions dealing with their native speaker status, and fifteen of the thirty-eight were found to be "tainted", the typical instance being that of a Japanese-American student who had been exposed to a great deal of Japanese in the home. These "tainted" subjects' tests were scored separately, then statistically compared through an analysis of variance with the results from the "pure" listeners. Amazingly, there was not only no significant difference in the performance of the two groups, but virtually no difference whatsoever. Thereafter, the two groups' scores were merged.

The listeners were presented with two samples of speech in each item of the test, the samples being either AB, BA, AA or BB in their relationship to the test categories. Thus, if the test was of *Yes*^{1} (statement) versus *Yes*^{2} (incomplete statement), the listeners were prepared (by means of a brief warm-up, utilizing categories related to, but not duplicates of the test categories) to answer either *Yes*^{1}, *Yes*^{2} or *Yes*^{2}, *Yes*^{1} or *Yes*^{1}, *Yes*^{1} or *Yes*^{2}, *Yes*^{2}. Each part of the test contained at least one example in which the two tokens were identical (i.e., an example edited into the form, say, *Yes*^{1}, *Yes*^{1}), and the listeners gave overwhelming evidence that they could hear identity and label it as such.

On the following page are reproduced the entire instructions (including the entire warm-up) for the listening test. Notice that the term INTONATION is not defined for the listeners. They defined it themselves through their choices on the test.

Directions for the Listening Test

In this experiment we are interested in learning how well you can discriminate between examples of American English INTONATION.

In each part of this experiment, you will hear words or phrases said in pairs. *Both* items in each pair will be spoken by the *same* speaker, and each pair will be preceded by that same speaker saying "Today is Monday" (to let you know what the speaker's voice sounds like).

The first item in each pair we will call "A", the second "B". Your task will be to listen to "A" and "B", and then to place them in the appropriate columns on the answer sheet. You may put "A" and "B" in different columns, or you may put them both in the same column, if they seem to you to belong in the same column.

For example, let's listen to two examples of the word "No", then place them in the appropriate columns. The two columns stand for "Sounds more like a question" and "Sounds more like an exclamation."

Question   Exclamation

Example 1)   ________   ________

Presumably you all placed "B" in the "Question" column, and "A" in the "Exclamation" column. Now let's try it again:

Question   Exclamation

Example 2)   ________   ________

Presumably you placed both "A" and "B" in the "Exclamation" column, because they both sounded much more like exclamations than like questions.

Now, unless you have any questions as to what you are to do, we will begin the experiment.

There were two grounds for our decision to play the "reference tone"-like phrase *Today is Monday*, as spoken by each speaker, before each test item for that speaker. The first was the experience from our pilot study, in which listeners became almost giddy from hearing many brief snatches of speech coming at them with no context whatsoever. The other was the suggestion by Martin Joos that perhaps one reason for the universality of phatic speech (e.g., "How do you do?") was the necessity for the listener to get some idea of the speaker's phonological system before crucial commincation took place. (Joos, 1948: 61-62) By having each speaker record the quite neutral statement *Today is Monday*, and by playing that utterance before each speech sample from that speaker, we were able to reduce the feeling of linguistic anomie on the part of the listener, without giving him too much of a headstart on the test items themselves. It should be remembered that there were twelve different voices to be listened to, and that the speakers were played in different order from one part of the test to the next. This variation, together with the two five-minute breaks (which extended the total testing session time to one hour and twenty minutes) successfully prevented the occurrence of any learning effects, as measured by both analysis of variance and binomial distribution tests.

Two test orders were used, with items essentially reversed within the major categories (i.e., syntactic and emotional items). The use of the same tests as above showed no ordering effects. As given in test order #1, the test items were the following:

Syntactic Categories

Were^{1} (*were they black* spoken as a question)	versus	*Were*^{3} (*were they black* spoken as a subjunctive)
Yes^{1}	versus	*Yes*^{4}
Yes$^{\text{(unemphatic)}}$	versus	*Yes*$^{\text{(emphatic)}}$
Were^{2} (incomplete, "or"-type question)	versus	*Were*^{3} (subjunctive)
Yes^{1}	versus	*Yes*^{2}

Ridiculous$^{\text{(unemphatic)}}$	versus	*Ridiculous*$^{\text{(emphatic)}}$
Were^{1}	versus	*Were*^{2}
Yes^{3}	versus	*Yes*^{4}

The reader will note that we follow Bierwisch (1966) in considering emphasis part of the syntactic component of a grammar. Stockwell (1959) already made this point.

Emotional Categories

Yes(agreeable)	versus	*Yes*(disagreeable)
Yes(calm)	versus	*Yes*(angry)
Yes(bored)	versus	*Yes*(interested)
Yes(belief)	versus	*Yes*(disbelief)
Yes(angry)	versus	*Yes*(contained anger)

Within each test category, the listeners heard the twelve speakers, in varying order. Each test item consisted of a speaker saying *Today is Monday*, then the two examples of his speech (e.g., Yes^1 versus Yes^4). The time lapses were varied slightly, to avoid establishing a montonous rhythm, but averaged 2.9 seconds between *Today is Monday* and the first test sample, 2.6 seconds between the first and second test samples, and 4.5 seconds between the end of the second test sample and the beginning of the next test item. Because an electronic editing device was not yet available in the laboratory, it was necessary to edit by tape copying. This produced some slight distortion on ten examples of the *were they black* items, where great precision was necessary in separating the test items from their contexts; but this distortion, manifested primarily in the form of an abrupt beginning to the phrase, did not interfere with the listeners' perception of a clearly-made contrast, as shown by the fact that on all the *were*^{1}/*were*^{2}/*were*^{3} contrasts, the worst communication scores were made by speakers whose utterances had no distortion, while the highest communication scores were made by both distorted and undistorted speakers. Similarly, in the case of slightly increased background noise in the recording of Speaker 6 and, to a much lesser extent, Speaker 2, listeners were able to hear the contrasts (when they had been made) above the noise. On more than 75 percent of the items, however, there was neither distortion nor excessive background noise.

Equipment and Measurement Procedures Utilized

Each speaker was seated in a large recording booth (IAC model 400ATR) in the UCLA Phonetics Laboratory, and was recorded on an Ampex console tape recorder which, at the 7 1/2 ips speed used, showed a response of ± 2 db from 50 to 12,000 Hz. The tape recorders used for editing had a frequency response of + 3 db from 70 to 10,000 Hz, and a signal/noise ratio of more than 40 db. The pitch meter used in the initial stages of fundamental frequency measurement (see below) was built in the Speech Transmission Laboratory, R.I.T., Stockholm, and has been described by Risberg (1962). The signals for fundamental frequency, amplitude, and wave-form were fed through three channels of a Siemens Oscillomink at a speed of 10 cm/sec. However, because of the weaknesses inherent in pitch meters (namely, a tendency for the pitch to "drop out" at sometimes crucial points in the curve, unless the meter's frequency curve is excessively "smoothed"), we adopted an additional procedure for measuring f_o. More than 350 narrow-band spectrograms were run off on a Kay Electric Sona-Graph sound spectrograph, modified and used as described in Ladefoged (1962b). These spectrograms were carefully measured on the best harmonic (ranging from the fourth to the eleventh), with the measurements constantly being compared to the f_o curves previously obtained from the pitch meter. The figures for f_o reported herein therefore represent a collation of the best data available for each utterance. Therefore, while cognizant of the problems of frequency measurement on the Sona-Graph (cf. Lindblom, 1962), we feel that our f_o figures are worthy of confidence.

For the purpose of providing illustrations for our data, we then measured each spectrogram on the fifth harmonic (reconstructing that harmonic from other harmonics where necessary), made a tracing of the fifth harmonic, and had the tracing photographically reduced, so that the spectrogram f_o curve would have exactly the same time scale as the amplitude and wave-form displays produced on the Oscillomink. Each illustration, therefore, shows a fundamental frequency curve collated from the best available data, superimposed upon time-matched Oscillomink displays of amplitude and wave-form. Dotted segments on some of the f_o curves indicate a lesser degree of confidence, because of the lessened voice amplitude at the beginnings and ends of utterances.

A further note upon equipment and procedures is necessary here. Most of the subjects heard the tape for the listening test from a Ferrograph type 5A tape recorder playing through a Sony SSA 777 monophonic amplifier-speaker system, while seated in a small room in the UCLA Phonetics Laboratory. However, some of the listeners in Order #2 had to take the test in a much larger toom, equipped with a somewhat inferior speaker system. Interestingly, they did not make significantly more wrong judgments, but they did fail to hear a difference somewhat more often than did those who heard the same tape under better listening conditions. This slightly larger number of neutralized judgments almost (but not quite) caused an order difference between the two groups (Order #1 versus Order #2) of listeners, as measured by both analysis of variance and binomial distribution tests.

Assumptions Underlying the Statistical Analysis

The basic statistical tool utilized was analysis of variance. This was supplemented by the use of the Newman-Kuels test (Winer, 1962: 80-85) for ranking items on the basis of the number of neutralized judgments, and an *an hoc* "Communication Score" (C-score). These procedures will be explained below.

In performing the analysis of variance, we assumed (as the null hypothesis) that if there were no communication taking place, then the listeners' responses would be random. We further assumed that random listener performance would manifest itself in the equal use of all four possible judgment categories (AB, BA, AA, BB), so that, in the null hypothesis, only one out of four judgments would be correct. Given the fifty listeners used in our test, random performance would yield an average of 12.5 correct answers on each test item. Underlying this assumption concerning the null hypothesis is the further assumption that, from beginning to end of the listening test, subjects were equally willing to make use of all four judgment categories. We can give three different kinds of evidence in favor of this latter assumption. First, there is the matter of those test items which were deliberately edited into the form AA or BB (see p. 41 above). Each test category contained at least one such deliberately neutralized example, and three categories (*Were*^{1}/*Were*^{3}, *Ridiculous*$^{\text{unemphatic}}$/*Ridiculous*$^{\text{emphatic}}$, and *Yes*^{3}/*Yes*^{4}) contained two such items. These sixteen deliberately neutralized items, multiplied by the fifty listeners, made a total of 800 possible neutralized judgments. As we said on p. 41, the listeners gave overwhelming evidence that they could hear identity and label it as such. On the eleven such items in the "Syntactic" part of the test, they correctly labeled as undifferentiated 523 out of 550 judgments. On the six "Emotional" items, they correctly labeled 246 out of 250, yielding a total of 769 out of 800.

In the case of the contrasting test items, we have evidence of all kinds of judgments: correct (p), incorrect (*p), and neutralized (either as AA or BB -- see p. 49). These judgments all exist in sufficient quantity to justify confidence in the listeners' willingness to use all four judgment categories. Furthermore, there was no significant decrease in the number of neutralized judgments from the beginning to the end of the test. On the basis of this evidence, we conclude that the hypothesized figure of one out of four judgments correct by chance is justified, and on this basis compute the results such that, on a particular test item (e.g., Speaker 1's *Yes*^{1}/*Yes*^{2}) 21 or more correct judgments would be significant at the .01 level, and 24 or more correct judgments at the .001 level.

If, however, the hypothesized chance figure is contested, and it is maintained that the two neutralized judgment categories (AA and BB) should be collapsed into one, reducing the number of judgment categories to three, and thereby raising the chance possibility of a correct answer to one in three, this would not greatly affect the results. On this basis it would require 26 or more correct answers to reach the .01 significance level, and 28 or more for a significance level of .001. Since none of the test categories show a difference in measurement between 26 versus 28 correct judgments, we need not report the 26-or-more figure, but will report only how many speakers in each test category satisfied the 21, 24, and 28-or-more levels, i.e., how many speakers manifested performance in making intonation contrasts which was significant at .01 and .001 (for a chance possibility of one in four correct judgments), and at the .001 significance level (for one in three judgments correct by chance). There is actually only one case (*Were*^{1}/*Were*^{3}) where there is a difference in the number of speakers satisfying p = 24+ versus the number satisfying the p = 28+ measurement level. This would tend to lead us some small distance toward an "all-or-none" interpretation of performance in communicating intonation contrasts. When a speaker makes a definite contrast between two intonation contours, an extremely significant number of listeners hear the contrast correctly. When the contrast is not made clearly, then lowering the desired significance level does not seem to help very much.

A Newman-Kuels *post-hoc* comparison test for significant analysis of variance was done on the test categories in respect to the number of neutralized responses (AA or BB judgments on items which were intended as AB or BA). According to this test, the categories group as follows:

Table II

Category	Number of neutralized judgments
Yes^3/*Yes*^{4}	195/600
Were^2/*Were*^{3}	134/600
Were^1/*Were*^{2}	125/600
Yes^1/*Yes*^{4}	111/600

Were^{1}/*Were*^{3}	110/600
Yes$^{\text{belief}}$/*Yes*$^{\text{disbelief}}$	103/600
Yes$^{\text{angry}}$/*Yes*$^{\text{contained}}$	100/600
Yes$^{\text{agreeable}}$/*Yes*$^{\text{disagreeable}}$	94/600

Yes^{1}/*Yes*^{2}	52/600
Ridiculous$^{\text{unemphatic}}$/*Ridiculous*$^{\text{emphatic}}$	39/600
Yes$^{\text{unemphatic}}$/*Yes*$^{\text{emphatic}}$	31/600
Yes$^{\text{calm}}$/*Yes*$^{\text{angry}}$	12/600

However, we feel that even though the *Yes*^{3}/*Yes*^{4} category is significantly different from the other categories, it would be incorrect to interpret this neutralization index as indicating that certain contrasts are intrinsically more difficult to hear. Let us contrast listener reactions to two speakers' utterances of *Yes*^{3}/*Yes*^{4}, the most "difficult" category. Speaker 12 was one of the two best performers in this category. In her case, 42 of the 50 listeners heard the contrast and labeled it correctly. Another seven heard it the wrong way around, and only one listener gave a neutralized judgment (labeling both utterances as more like questions). But in the case of Speaker 6, the results are quite different. There, only seven listeners labeled the contrast correctly, and another seven heard it the wrong way around, while 36 gave neutralized judgments (28 labeling both items as *Yes*^{3}, eight labeling them both as *Yes*^{4}). When we examine the stimuli, the reasons are immediately obvious. Speaker 12's utterances (illustrated in Fig. 4.7, p. 59 below) contrast markedly. Her *Yes*^{3} shows a strong scoop at the nucleus of the contour, with a moderately high terminal rise, while her *Yes*^{4} has no downward scoop, and takes off quite early toward a very high and steep terminal rise. However, Speaker 6 has virtually no contrast between her utterances. Both her *Yes*^{3} and *Yes*^{4} have a definite scoop at the nucleus, and the terminal rise of the *Yes*^{4} is only a little higher than that on the *Yes*^{3}. The conclusion seems obvious: when a definite intonation contrast is made (even in a somewhat esoteric category such as *Yes*^{3}/*Yes*^{4}) the listeners hear it. When it is not made, they do not hear it. Thus it is not true that some of the contrasts dealt with in this study are intrinsically more difficult to hear. Rather, it is the case that a smaller number of speakers normally (i.e., without "coaching") produce those contrasts clearly enough for them to be unambiguously heard by listeners. (For further discussion, please see p. 55 below)

Following the above interpretation, we would conclude that the Newman-Kuels ranking of difficulty according to the number of neutralized judgments does not cast any light on the nature of the stimuli which is not already supplied by our other tests. There is, however, one exception to this statement. The *Yes*^{1}/*Yes*^{4} category (see pp. 52-54 below) was one in which the speakers performed very well, according to our analysis of variance and C-score measurements. Yet the listeners gave neutralized judgments on 111/600 test items, and incorrect judgments (BA instead of AB) on only 14 items, with correct judgments on 475/600 items. This extremely low number of incorrect judgments, combined with the relatively large number of neutralized judgments, would seem to be one more bit of evidence for that ambiguity concerning the very nature of questions and statements which is so perfectly expressed in Uldall's famous subtitle: "Are You Asking Me or Telling Me?" (Uldall, 1962).

In the measurement of speaker performance on intonation, neutralized listener judgments are not as serious, in our view, as incorrect judgments. We believe that this interpretation follows from the observation of Uldall which we quoted on p. 1 above, to the effect that "the same kind of information is carried by several systems all present at all times: pitch, voice-quality, tempo, gesture, facial expression...." Although our data in the next two chapters suggest some limitations upon this viewpoint, it is generally true that, if the intonation contour is not in itself misleading, the desired syntactic and/or emotional information may be communicated by other means, ranging from word order to gesture. However, a misleading intonation could seriously interfere (sometimes only temporarily) with the desired communication. For this reason, it would seem reasonable, in rating speaker performance, to "punish" speakers for listener judgments which were incorrect. Furthermore, a measurement scale which comprehended both correct and incorrect judgments might also serve as a means (given the above assumptions concerning the communication of information through intonation) for separating the really gifted communicators from the mass of those whose performance satisfied even a high significance level. Therefore, we utilized an *ad hoc* "Communication Score" (C-score), which is derived by subtracting the incorrect judgments from the correct judgments, and multiplying the result by two, so that C = 2(p-*p). This yields a measurement scale running from 100 to -100. Using a significance level of .01 on a two-tailed test, the chance area within the larger range is ±10.72 for each category (all speakers). For each speaker within a particular category, the chance area is ±38.60. Thus, our lowest recorded mean C-score (*Yes*^{3}/*Yes*^{4}, p. 57) is still well above the level of chance performance. Similarly, on the *Yes*^{1}/*Yes*^{4} contrast, Speaker 11's "worst" performance (pp. 54-55), with its C-score of 40, is still slightly above the chance level. All of the "best" examples reproduced on the following pages easily exceeded the chance level, and scores above 90 were surprisingly frequent. We suspected that a zero score would indicate complete neutralization in the stimuli (i.e., the speaker did not produce anything which could be called a contrast). Such a zero score was recorded for the *Yes*^{3}/*Yes*^{4} of Speaker 6 which was discussed on p. 48 above. Scores with a large minus value were considered as tending toward the anomalous, and the two worst (a -84 and a -62) are illustrated and discussed in Chapter 5.

For each contrast which is illustrated and discussed in the next two chapters, we will show how many of the twelve speakers performed at three different significance levels (see p. 48 above). We will also give the mean and median C-scores for the entire group of speakers (the median being the more significant figure, since one anomalous test item from one speaker could produce a minus C-score which would drag the mean of the C-score down rather considerably). We have chosen the two "best" and one "worst" test items for illustration below on the basis of C-scores, and have given those scores in the captions. In the case of the two contrasts which had very anomalous "worst" items, we have added illustrations of a "typical worst" test item, i.e., a pair of stimuli more typical of those receiving low C-scores in that test category. In this way we hope to clarify for the reader what kinds of contrasts were effective, less effective, and anomalous.

Chapter 4: Syntactic Aspects of American English Intonation

In this chapter are discussed the experimental results for the eight syntactic categories which were explained in detail on pp. 33-39 above. A technical discussion of the illustrations will be found on pp. 46-47. While our findings for the "unemphatic"/"emphatic" contrasts will come as nc surprise, we feel that the results for the contrasts involving the test phrase *were they black*, and for some of the contrasts using the test word *yes*, will be of interest to those working in the field of English intonation.

Discussion of Syntactic Test Categories

Yes^{1} (simple statement)/*Yes*^{4} (repetitive question)

Mean C-score (all twelve speakers): 76.83

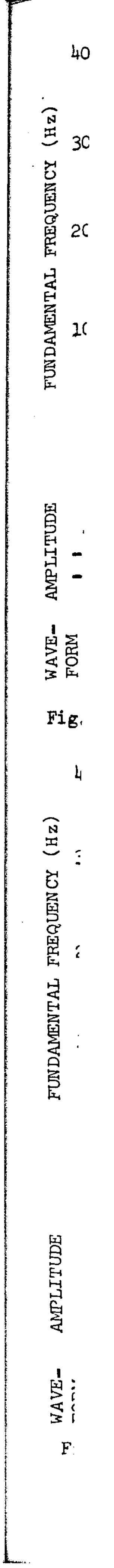
Median C-score: 90

Number of speakers for whom there were

21 or more correct judgments:	12/12
24 " " " "	10/12
28 " " " "	10/12

The dominant cue (cf. Figs. 4.1-2) in helping the listener to distinguish *Yes*^{4} (both here and in contrast with *Yes*$^{3}_{4}$ (continuative)) is the existence of a high, steep terminal rise on *Yes*^{4}. Looking at *Yes*^{1}, we immediately notice that it is not at all necessary for a contour to have a terminal fall in order for it to be clearly perceived as a statement, a fact previously noted by Uldall (1962) and Hadding-Koch and Studdert-Kennedy (1964). This is shown also in the very good (C = 90) performance of Speaker 3, who also had a slight terminal rise on his *Yes*^{1} contour.

A successful alternate gesture for *Yes*^{4} was manifested by Speaker 7, with a high rise from a precontour of 120 Hz to a nucleus at 200 Hz, then a fall to a turning point at 125 Hz, and a slight terminal rise to 135 Hz. If the usual *Yes*^{4} can be interpreted as "did you really say *yes*?" then Speaker 7's *Yes*^{4} might be interpreted as "You don't mean to say *yes*?" In



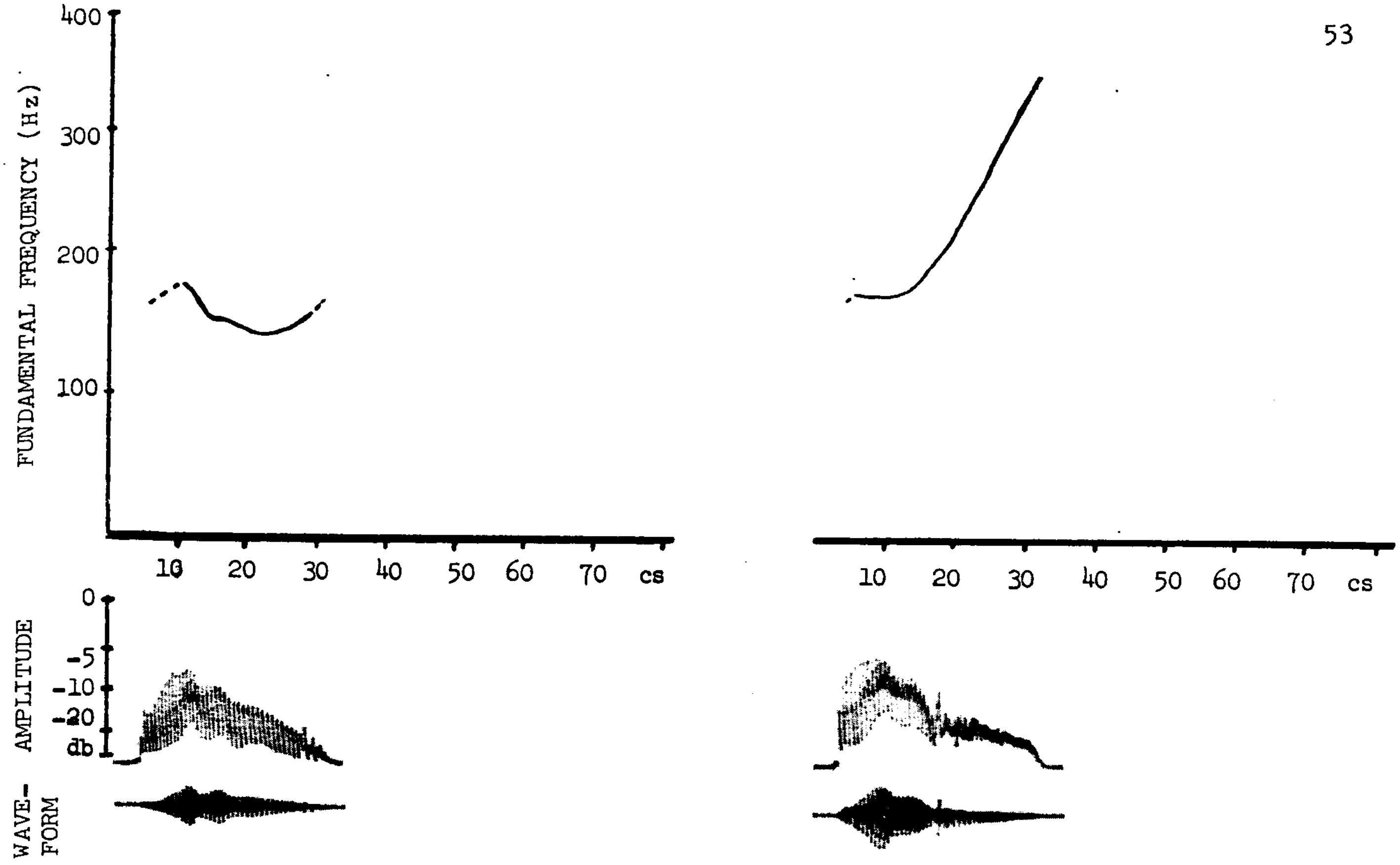


Fig. 4.1 Speaker 12 Yes^1 C = 96 Speaker 12 Yes^4

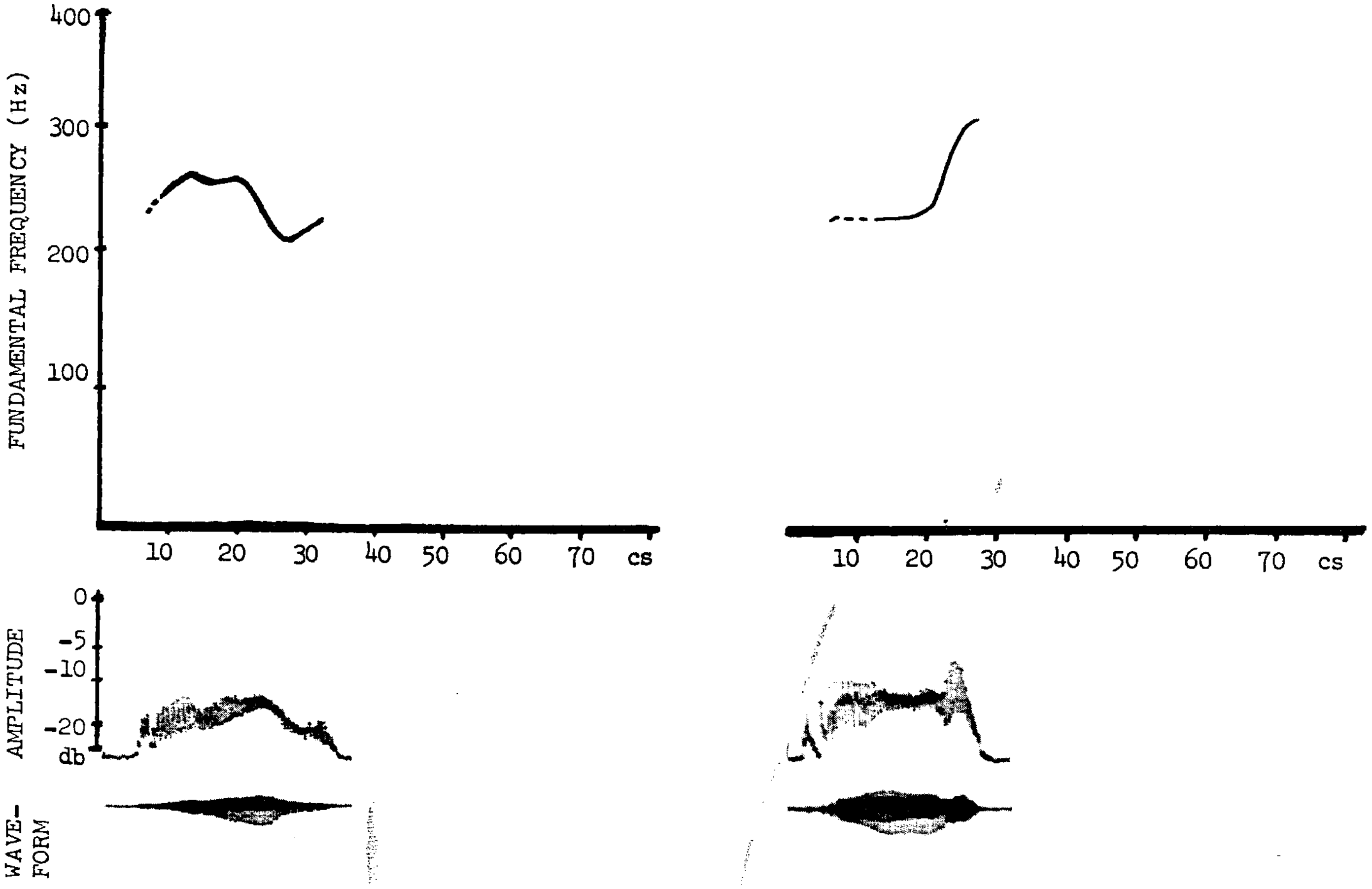


Fig. 4.2 Speaker 4 Yes^1 C = 96 Speaker 4 Yes^4

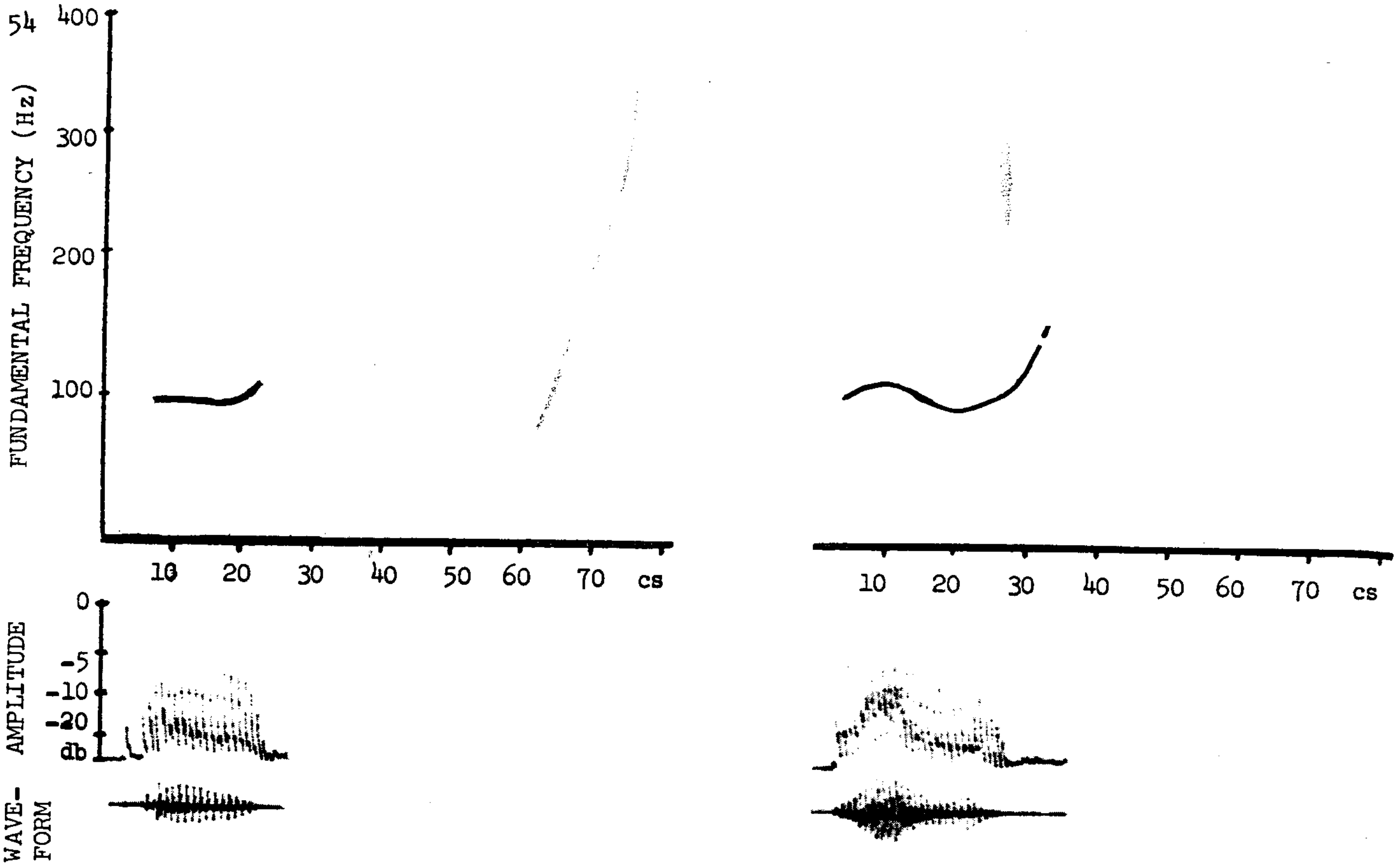


Fig. 4.3 Speaker 11 Yes^1 C = 40 Speaker 11 Yes^4

contrast with his *Yes*^{1}, which had a moderate nuclear rise from 160 to 180 Hz and a terminal fall to 90 Hz, Speaker 7's alternate *Yes*^{4} yielded a C-score of 92.

We have already noted that, with the exception of a relatively high number of neutralized judgments, the *Yes*^{1}/*Yes*^{4} category was one in which the speakers performed very well. A further indication of this excellent performance can be seen in the fact that our "worst" example (Speaker 11, Figure 4.3) achieved an above-chance score of C = 40. We interpret his failure to do better as stemming from two causes. First, there is only a moderate rise to the end of his *Yes*^{4} contour. This caused 27 of the 50 listeners to categorize both his *Yes*^{1} and *Yes*^{4} as statements. Also complicating the interpretation of his *Yes*^{4} is the definitely scooped nucleus, with the scooped portion of the contour having a high amplitude. Such a scooping to the nucleus is far more typical of the preferred *Yes*^{3} contour, as can be seen not only from the examples of *Yes*^{3}/*Yes*^{4} illustrated below (Figures 4.7-8), but also from the fact that Speaker 11 did rather poorly on his *Yes*^{3}/*Yes*^{4} contrast, with a C-score of 26, largely because 20 of the listeners interpreted both stimuli as *Yes*^{3}.

The second-worst performer in this category was Speaker 8 (C = 44), whose *Yes*^{1} had a fairly high terminal rise, causing 23 listeners to

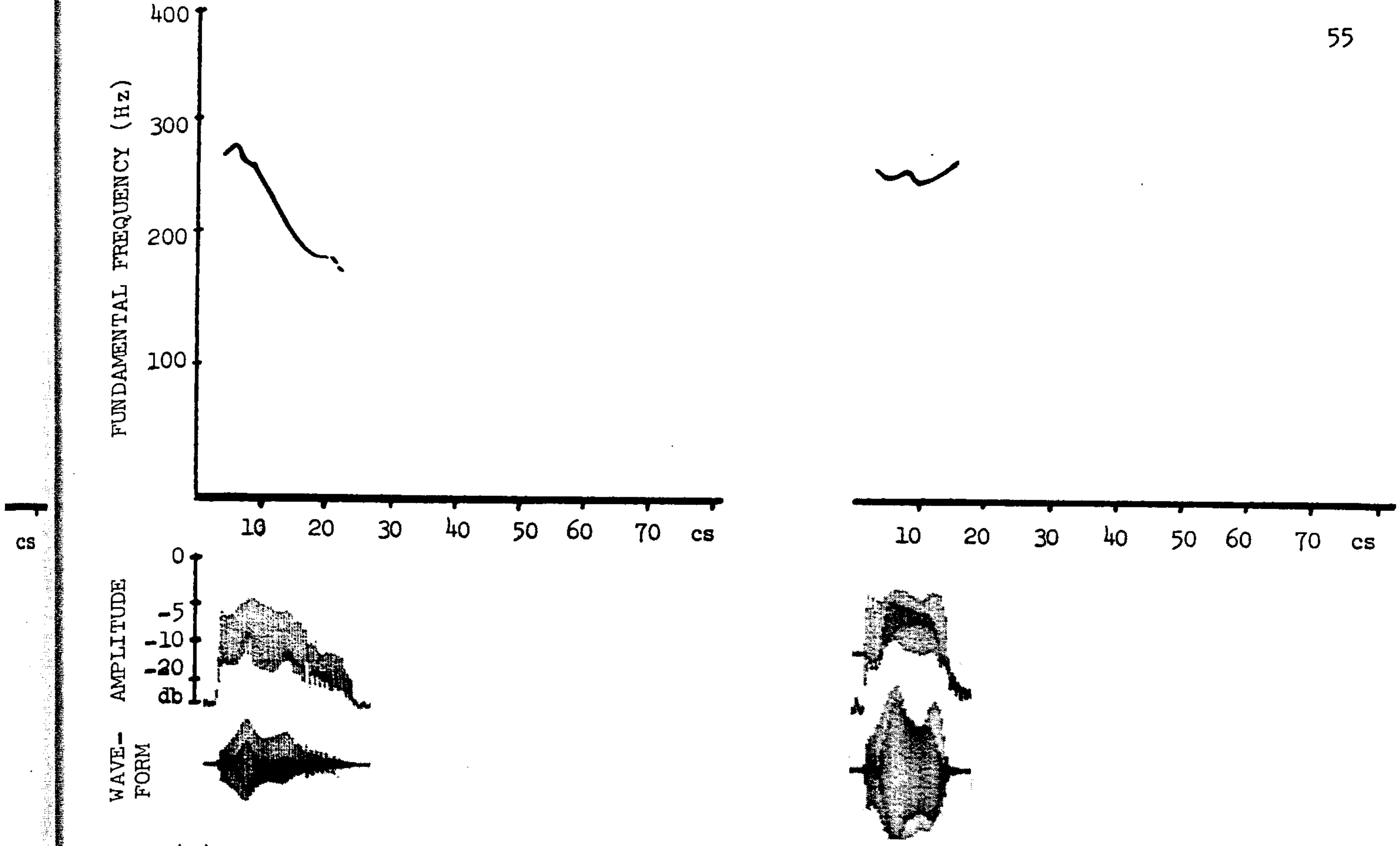


Fig. 4.4 Speaker 2 Yes^1 C = 96 Speaker 2 Yes^2

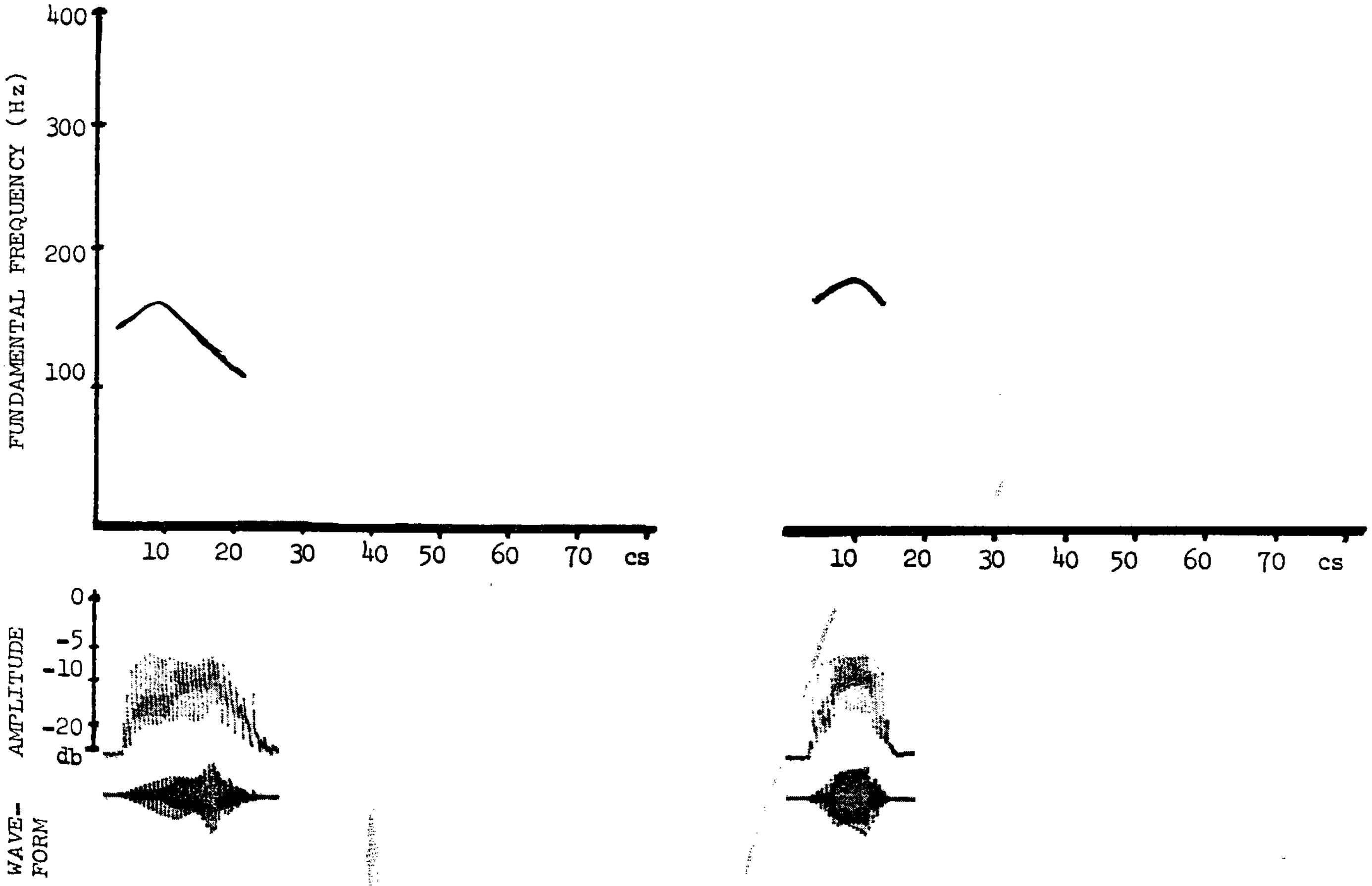


Fig. 4.5 Speaker 9 Yes^1 C = 96 Speaker 9 Yes^2

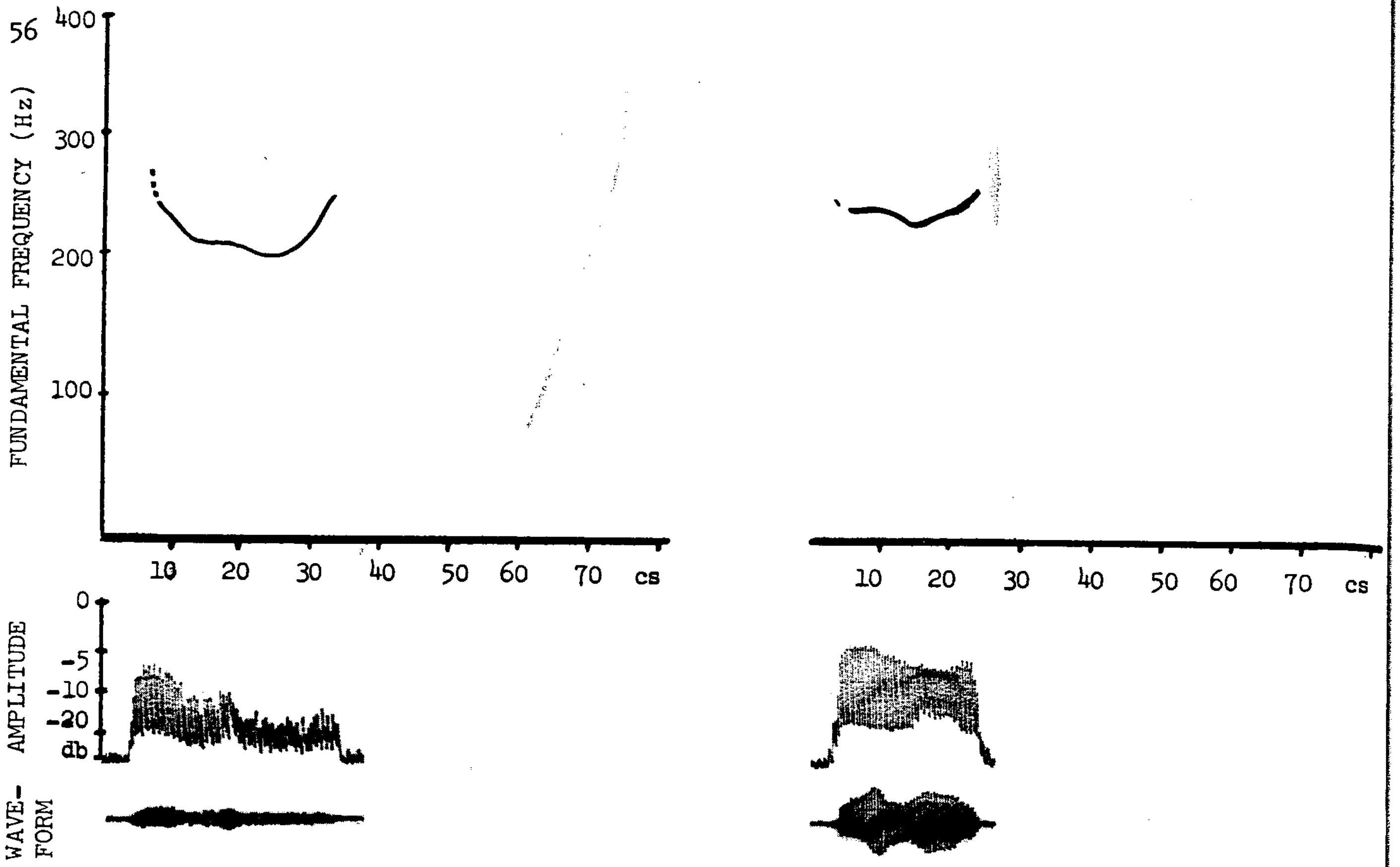


Fig. 4.6 Speaker 6 Yes^1 C = 70 Speaker 6 Yes^2

categorize both stimuli as *Yes*^{4}.

Yes^1(simple statement)/*Yes*^{2}(incomplete statement)

Mean C-score: 85

Median C-score: 85

Number of speakers for whom there were

21 or more correct judgments: 12/12

28 or more correct judgments: 12/12

As the above figures show, all of the speakers were very successful in communicating this contrast between a normal statement and an incomplete statement. As in the examples in Figures 4.4-5, the *Yes*^{2} contour is generally briefer, at a relatively high f_o, and with a relatively constricted frequency range. Its end point, while not perfectly level, has no marked terminal rise or fall. The best contrast is with a *Yes*^{1} which is relatively long, and has a marked terminal fall. But this is not necessarily the case. Speaker 8 (C = 92) had a high, brief (14 centiseconds), narrow ranged *Yes*^{2} contrasting with a longer 925 cs) *Yes*^{1} which had a sharply scooped nucleus and a relatively high terminal rise. One suspects that in this case duration may have been the dominant cue.

Yes^3(continuative)/*Yes*^{4}(repetitive question)

Mean C-score: 23.50

Median C-score: 26

Number of speakers for whom there were

21 or more correct judgments: 6/12

28 or more correct judgments: 6/12

By all standards of measurement, this was the most difficult of the test categories. On no other category did so few speakers achieve significant levels of performance (as determined by analysis of variance), and the C-scores here are much lower than for any other category, with only 4/12 speakers satisfying the C-score significance level. No other category had more than one speaker with a minus C-score. Here, one speaker scored a zero (see p. 49), and four speakers were ranged from -4 to a low of -30 (Speaker 5, illustrated on p. 83). We said on page 50 that we felt that this difficulty was a function of how few speakers were experienced in making a clear contrast between these two contours, since, in the case of Speaker 12, her C-score of 70 resulted from 42 listeners' labelling her contrast correctly, seven incorrectly, with one neutralized judgment -- performance which would be significant by any scale of measurement. But we still believe that this category was not difficult to *hear*; it may be true that it was difficult for the listeners to *conceptualize*. One indication of this conceptual difficulty can be inferred from the directions supplied for the listening test. On most of the categories, these directions were extremely simple. For example, on the *Yes*^{1}/*Yes*^{4} test, the listeners were given the printed directions: "In this part you will hear some pairs of the word *yes*. One item in each pair may sound more like a statement. The other may sound more like a question." However, anticipating some listener difficulty in conceptualizing *Yes*^{3}, we resorted to the following directions for *Yes*^{3}/*Yes*^{4}:

> In this part you will hear some pairs of the word *yes*. One item in each pair may sound like a question. The other may sound like a request for another speaker to continue, i.e., as though one were saying "I am listening to you ... please go on with what you were saying."

This was as explicit as we dared be, for fear of biasing the listeners, and it is possible that our directions confused as much as they enlightened, and thereby kept the best speakers from achieving C-scores as high as those for the best speakers in the other test categories. It is impossible to say. However, we might point out that the test categories involving the subjunctive *Were*^{3} (a very esoteric item in contemporary American English) had entailed not only a test item which may have been difficult to conceptualize, but also somewhat complicated directions, yet, particularly in the case of *Were*^{1}(question)/*Were*^{3}(subjunctive), better performances were recorded.

Looking at the stimuli (Figures 4.7-9), we see that the dominant cue for distinguishing *Yes*^{4}(repetitive question) is once again the height and steepness of the terminal rise on the *Yes*^{4} contour. The best contrast is with a *Yes*^{3}(continuative) which has a scooped nucleus and a moderate terminal rise. In this respect, the case of Speaker 5 is extremely interesting. He had a slight scoop on both the *Yes*^{3} and *Yes*^{4} contours, and his *Yes*^{3} had the slightly higher terminal rise (170 Hz as opposed to 150 Hz on *Yes*^{4}). As a result, only 10/50 listeners labelled the contrast the way he had intended it, while 25 heard it the wrong way around, producing a C-score of -30. Since a majority of the listeners heard Speaker 5's *Yes*^{3} as a *Yes*^{4}, we might say that in this case the height of the terminal rise outweighed the scoop as a perceptual cue. But what if the scoop had been sharper? Along the same line, Speaker 4 also had a slightly scooped ^{3} with a terminal rise (to 350 Hz) higher than that on her *Yes*^{4} (315 Hz). Again, this resulted in a very poor score (C = -10), providing further support for the importance of a high terminal rise as a distinguishing mark for the *Yes*^{4} contour.

A fairly successful alternate gesture was made by Speaker 10, who had a moderately long (28 cs), strongly scooped *Yes*^{3} with a fairly high terminal rise (nucleus at 130 Hz, turning point at 140 Hz, end point at 280 Hz), which contrasted with a longer (42 cs) "exclamatory" *Yes*^{4} of the type described above for Speaker 7, with a very high nuclear peak (385 Hz) followed by a turning point at 170 Hz, and a slight terminal rise to 200 Hz. This contrast produced a C-score of 62, as compared with scores of 70 for the two "best" speakers.

Were^1/*Were*^{3} (*were they black* as question/subjunctive)

Mean C-score: 66.33

Median C-score: 80

Number of speakers for whom there were

21 or more correct judgments:	11/12
24 or more correct judgments:	11/12
28 or more correct judgments:	9/12

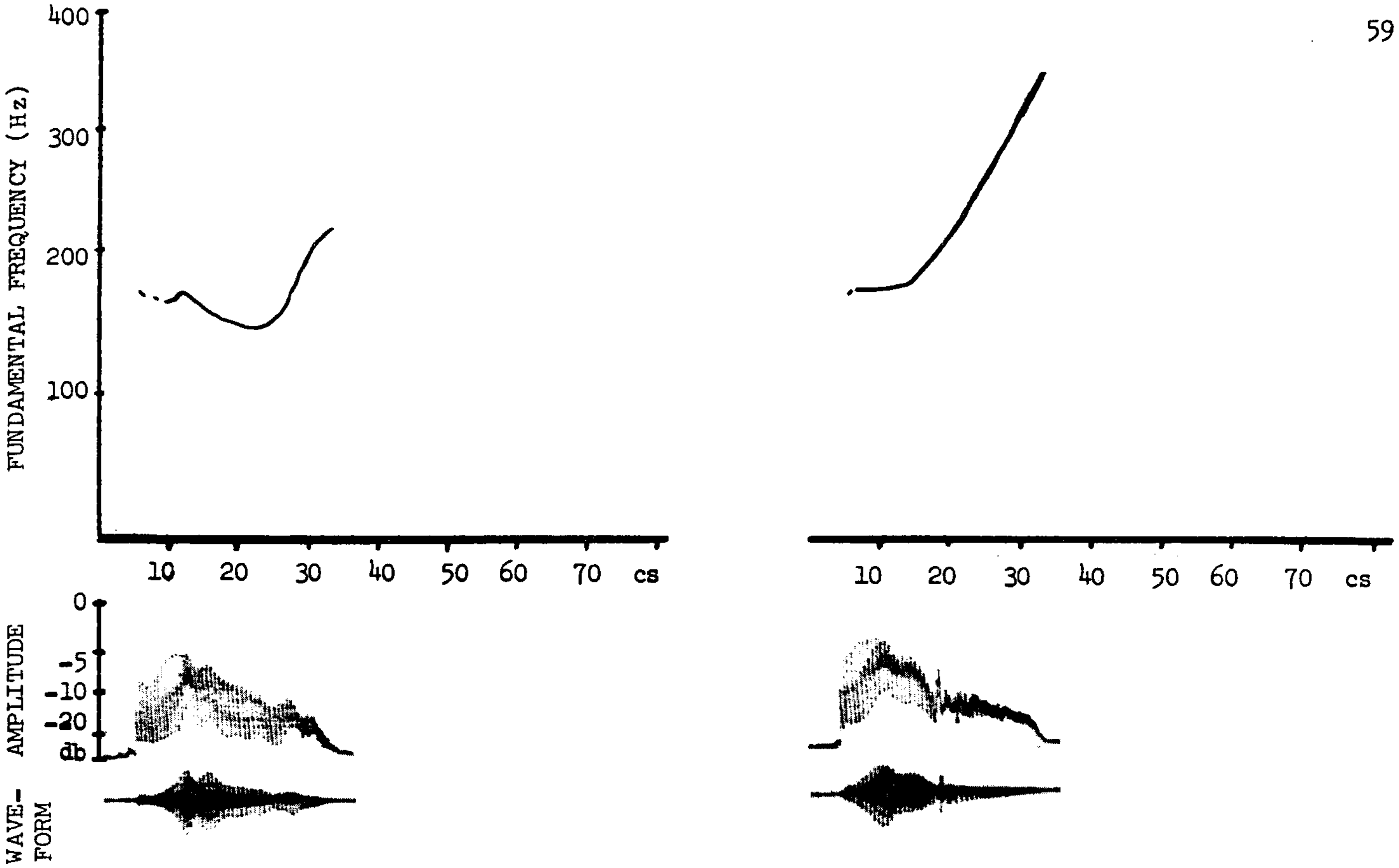


Fig. 4.7 Speaker 12 Yes^3 C = 70 Speaker 12 Yes^4

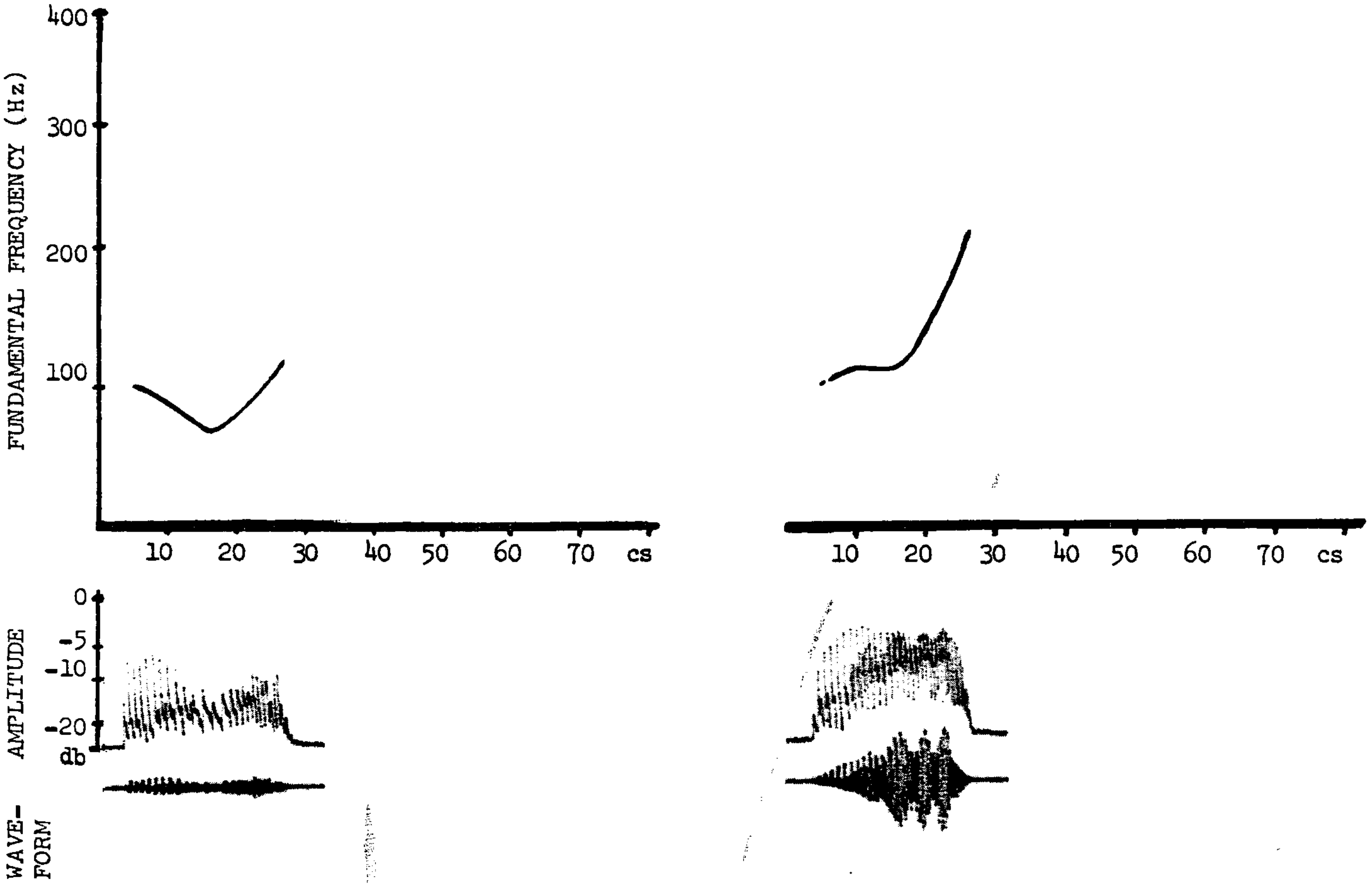


Fig. 4.8 Speaker 9 Yes^3 C = 70 Speaker 9 Yes^4

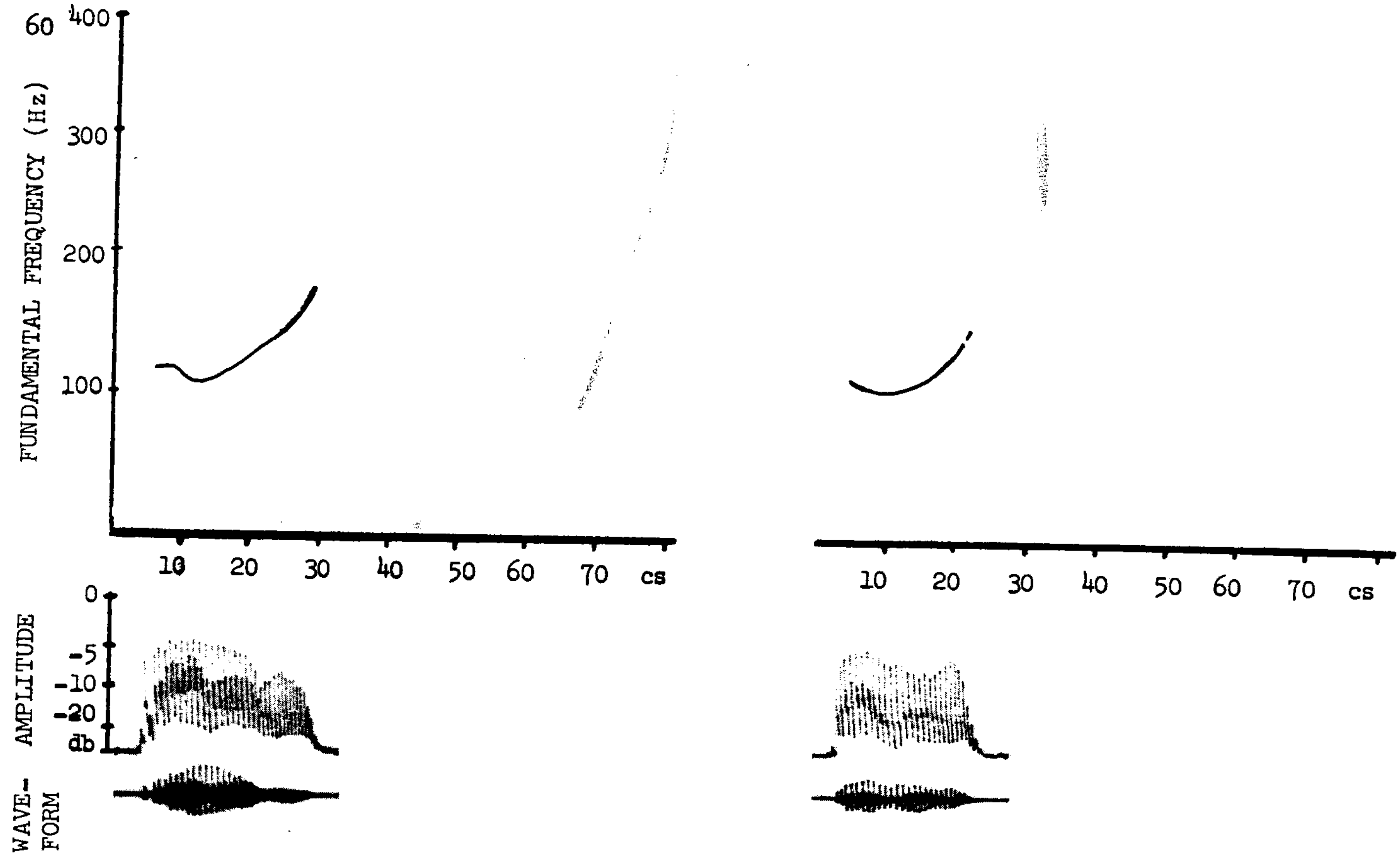


Fig. 4.9 Speaker 5 Yes^3 C = -30 Speaker 5 Yes^4

Concerning the nature of the stimuli, we should remind the reader that *Were*^{1} is a "normal" question, as opposed to the "repetitive" question we have previously seen in the case of *Yes*^{4}. We might expect that such a "normal" question, aided by "normal" question word order (see pp. 47-48, 50 above), could be signalled effectively by a less dramatic terminal rise than that required for effective performance with *Yes*^{4}. Although the two "best" speakers' utterances (Figures 4.10-11) are characterized by the same kind of high, steep terminal rise seen earlier for *Yes*^{4}, it was possible to achieve good communication with a lesser rise. Speaker 5, for example, had a *Were*^{3} with nuclear peak similar to that seen for Speaker 2's *Were*^{3}, preceded by a slightly falling precontour, and followed by a level terminal section. He contrasted this with a *Were*^{1} which had a relatively level precontour (115-125-115 Hz), a slight dip from 115-110 Hz at the nucleus, and a moderate terminal rise to 140 Hz. This contrast yielded a C-score of 84, with 45/50 listeners correctly categorizing the stimuli, and only three hearing it the wrong way around.

More problematical is the matter of *Were*^{3}, deliberately included because of its esoteric nature (see p. 50 above). Again anticipating conceptual difficulties, we resorted to elaborate directions for the listeners:

> In this part you will hear some pairs of the phrase *were they black*. One item in each pair may sound like a question. The other may sound like a somewhat old-fashioned use of the subjunctive, meaning "*If* they were black ..."

Again, it is possible that these directions confused as much as they enlightened (while administering the listening tests, we manfully ignored anguished cries of "What's a subjunctive?"). However, the extremely high communication scores for the best speakers, and the high over-all scores indicate that the listeners soon related the stimuli to appropriate categories in their underlying linguistic competence.

Examination of the best test items shows that the subjunctive *Were*^{3} is indicated by more than one factor. Immediately noticeable is the high nuclear peak on the vowel of *black*, and the non-finitive terminal segment (either slightly rising, sustained, or slightly falling). However, the *were* segment is also important, being generally higher, longer, and pronounced with greater amplitude than the *were* segment in *Were*^{1}. An extreme case of this pointing of the *were* segment can be seen in Speaker 9's *Were*^{3}, which assumed the shape

```
          re
         e
                              a    ck
        w       th                a
                   e      bl
                    y
```

and which, when contrasted with a *Were*^{1} contour very similar in shape to that of Speaker 6 (Fig. 4.11), resulted in a very high C-score of 90.

On the level of performance, there were some problems with *Were*^{3}. Three of the speakers stumbled over it, caught their error, and repeated it essentially correctly. On the level of underlying competence, it would be tempting to say that Speaker 3 (Fig. 4.12) lacked the *Were*^{1}/*Were*^{3} contrast, but this would be too simple a view, in the light of his very good performance on this contrast *as a listener*. For further discussion of this point, see the section dealing with competence and performance in Chapter 6. Speaker 3's very low score resulted from an extraordinarily high number of neutralized judgments. Of the 50 listeners, 40 categorized both his *Were*^{1} and *Were*^{3} as questions, and one listener labelled them both subjunctives. Speaker 3's *Were*^{3} had a very slight pointing of the *were* segment, but this was inadequate for distinguishing a subjunctive meaning, in view of the lack of any nuclear peak on the *black* segment.

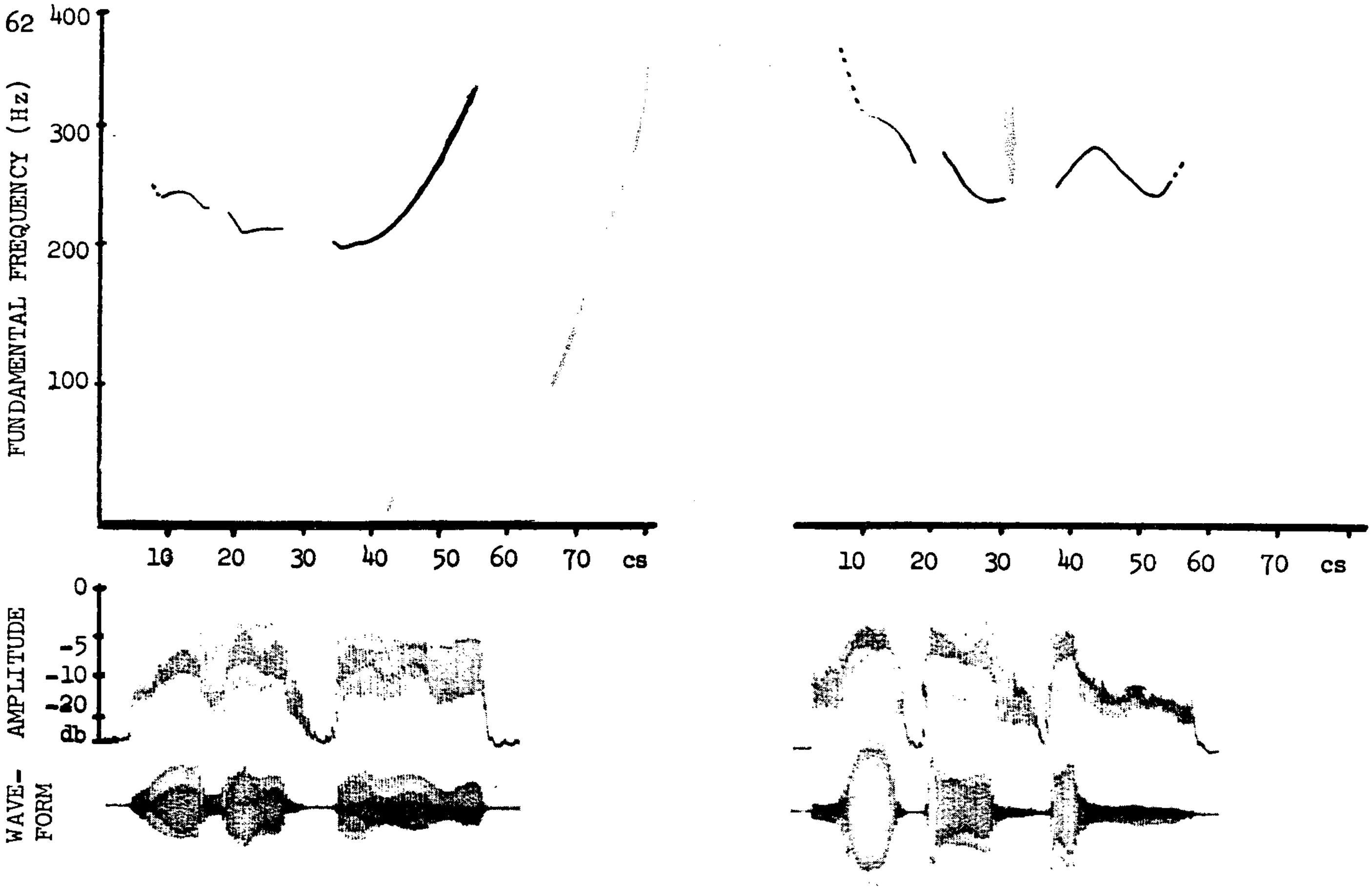


Fig. 4.10 Speaker 2 Were^{1} C = 94 Speaker 2 Were^{3}

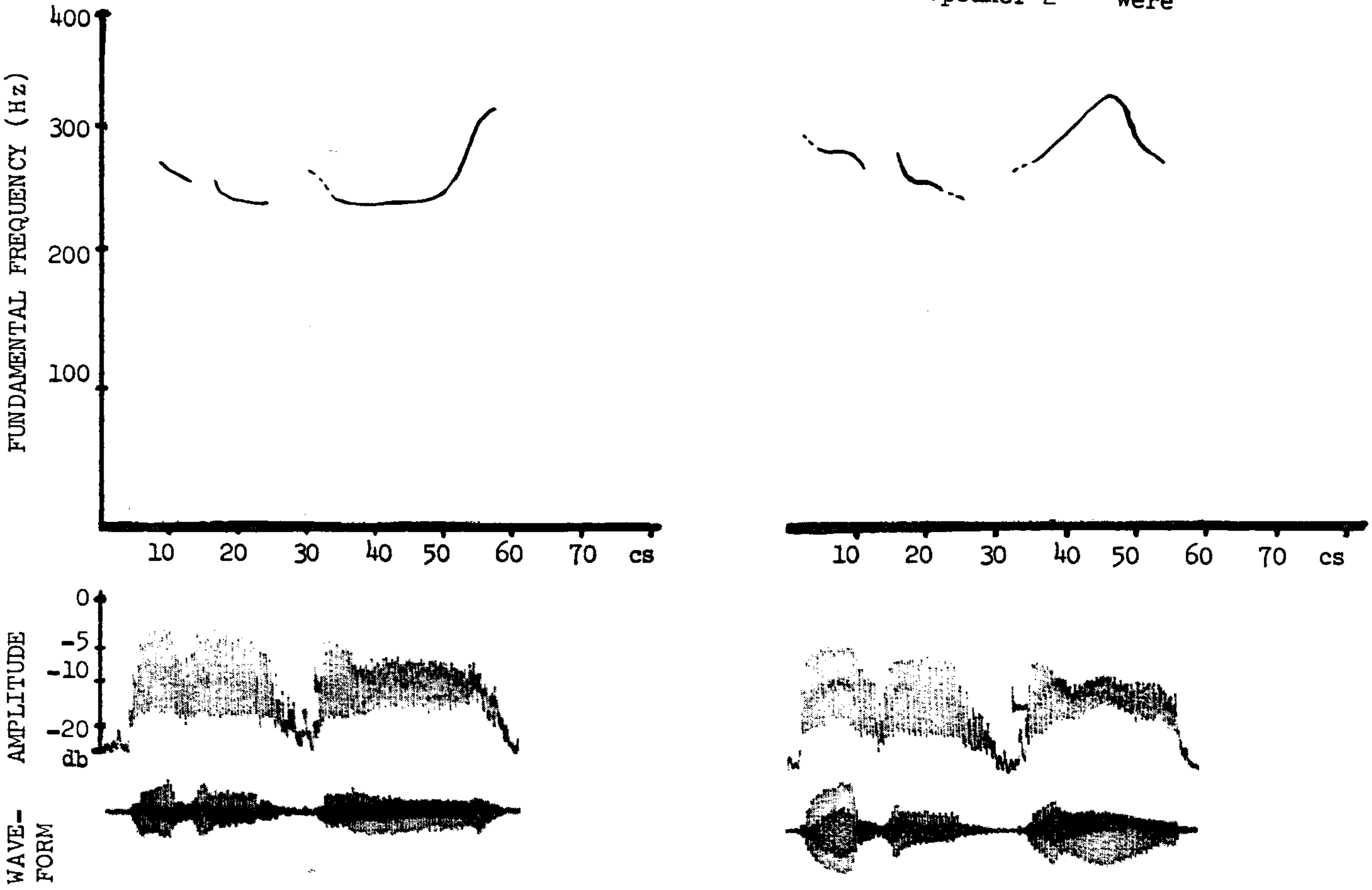


Fig. 4.11 Speaker 6 Were^{1} C = 92 Speaker 6 Were^{3}

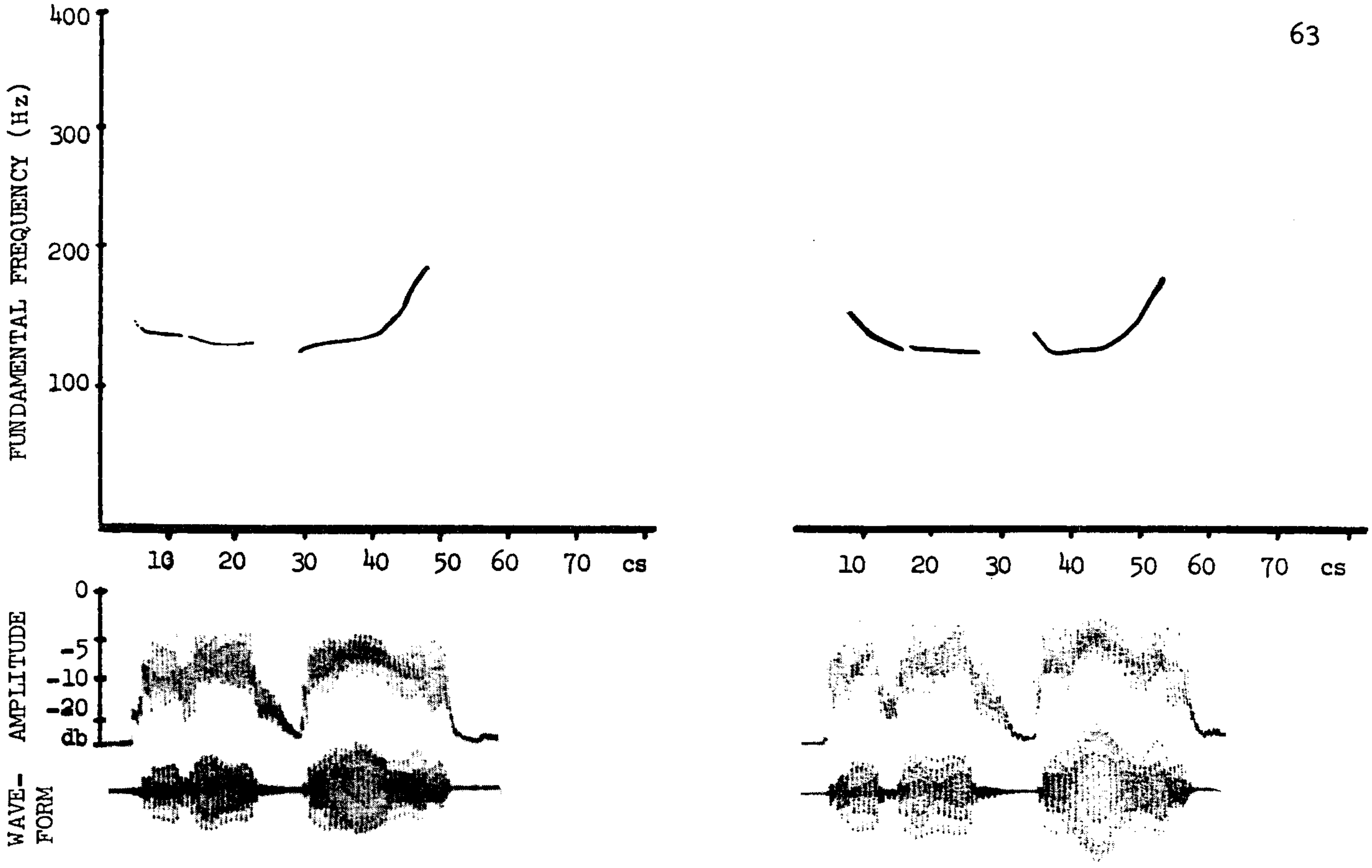


Fig. 4.12 Speaker 3 Were^1 C = -6 Speaker 3 Were^3

Were^1/*Were*^{2} (*were they black* as question/incomplete question)

Mean C-score: 49.67

Median C-score: 55

Number of speakers for whom there were

21 or more correct judgments: 10/12

24 or more correct judgments: 9/12

28 or more correct judgments: 9/12

This contrast between a "normal" question contour and the first half of an incomplete, *or*-type question can be viewed as a much more difficult case of the complete/incomplete contrast seen earlier with *Yes*^{1}/*Yes*^{2}. The difficulty stems from at least two sources. First, there is the loading of expectations toward a "normal" question interpretation of this word order, as discussed on p. 50 above. Second, this test phrase, consisting of three syllables containing several consonants, is simply not as elastic in terms of duration (assuming normal, unexaggerated speech) as the test word *yes*. Thus, differences in duration, which we found to be very important in distinguishing *Yes*^{1} from *Yes*^{2}, were in this case effectively eliminated as perceptual cues. This left only differences in end-of-phrase intonation as major cue. We have already noted (pp. 33-34 above) the doubts of both Hultzén and Bolinger that such a contrast (presumably manifested as rising versus sustained terminal contour segments) is normally made in American English. However, the great majority of the speakers did succeed in making a contrast between these two contours, as can be seen from the fact that nine of the twelve speakers performed at a very significant level (as measured by analysis of variance) and, on the C-score measurement, four speakers had C-scores of 70 or higher, and ten had C-scores of 38 or higher (see p. 51 for an explanation of the assumptions underlying this measurement device). Further more, they all effected the contrast in one of the two ways shown in Figs. 4.13-14. In other words, they had a slight terminal rise on *Were*^{1}(the question) contrasting with an essentially sustained terminal segment on the incomplete question *Were*^{2} (cf. Speaker 5, Fig. 4.13), or they contrasted a marked terminal rise on *Were*^{1} with a more modest rise on *Were*^{2} (cf. Speaker 8, Fig. 4.14). None of the speakers contrasted a marked terminal rise (as on Speaker 8's *Were*^{1}) with a sustained terminal segment (as on Speaker 5's *Were*^{2}). We suspect that such a definite contrast in end-of-phrase intonation would have yielded an extremely high C-score.

There is also a greater significance to precontour variations in this contrast than in the others, but those variations are so subtle, both in their physical characteristics and in their influence upon listener judgments, that it is difficult to discuss them in any quantitative form. We can only observe that the typical "best" *Were*^{2} (incomplete question) contour has a precontour which starts rather higher than the precontour for the *Were*^{1}(question) samples. This precontour usually falls somewhat, but it may remain essentially level. The higher beginning to the *Were*^{2} contour imparts a kind of "tentative", non-finitive air to *Were*^{2} which we believe aids in differentiating it from *Were*^{1}. It also makes whatever terminal rise exists less marked in nature. This relationship of the earlier portion of the contour to the terminal portion (cf. Hadding-Koch and Studdert-Kennedy, 1964 and Studdert-Kennedy and Hadding-Koch, 1969, forthcoming) can be seen operating negatively in the performance of Speaker 1 (Fig. 4.15). His *Were*^{1} contour is far more typical of the better examples of *Were*^{2}, and we believe that it was only the high precontour on his *Were*^{2} (reducing its similarity to the better examples of *Were*^{1}) which prevented more "wrong way around" judgments, and thereby kept him from having an even greater minus C-score.

Were^2/*Were*^{3} (*were they black* as incomplete question/subjunctive)

Mean C-score: 43.83

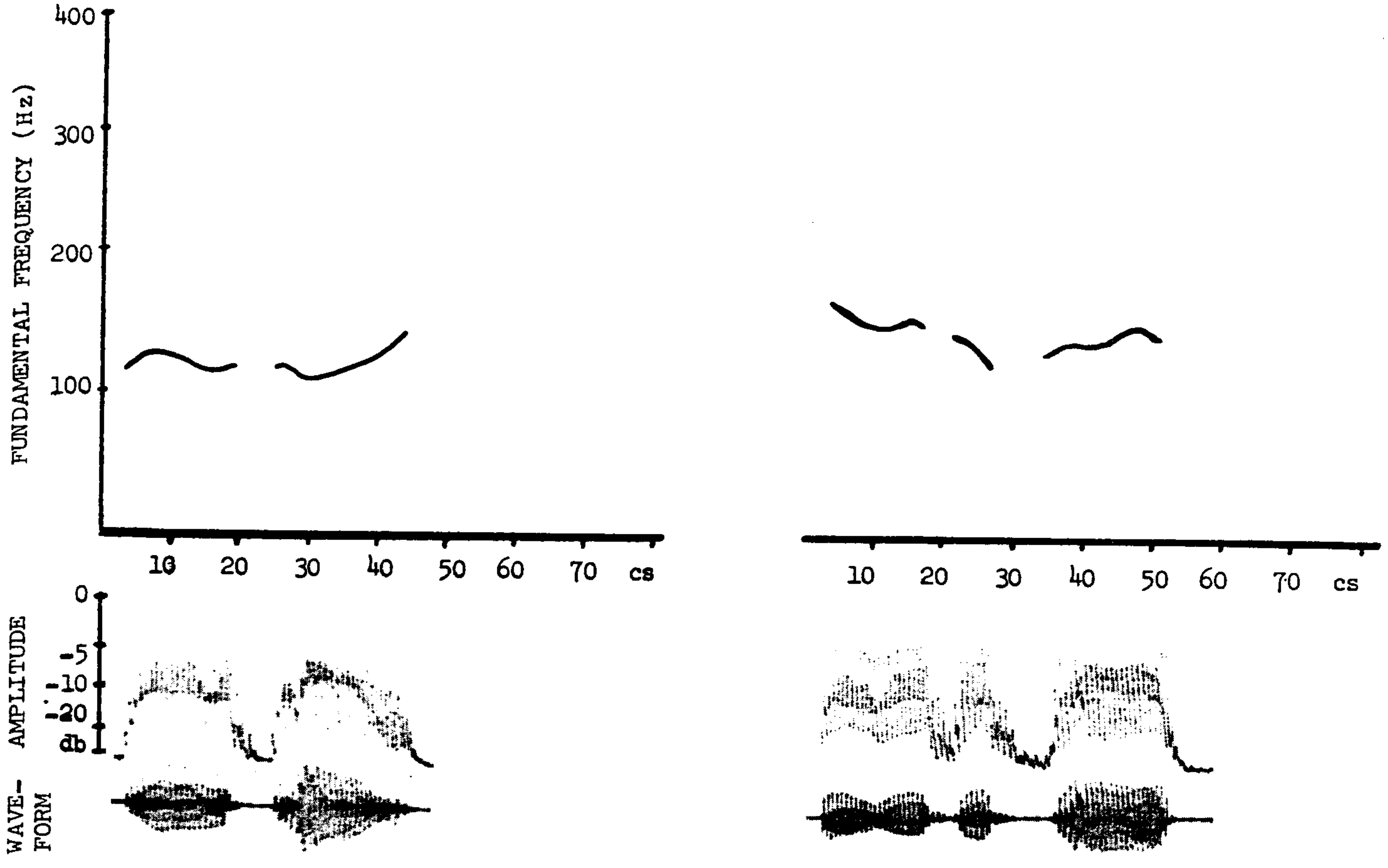


Fig. 4.13 Speaker 5 Were^1

C = 76

Speaker 5 Were^2

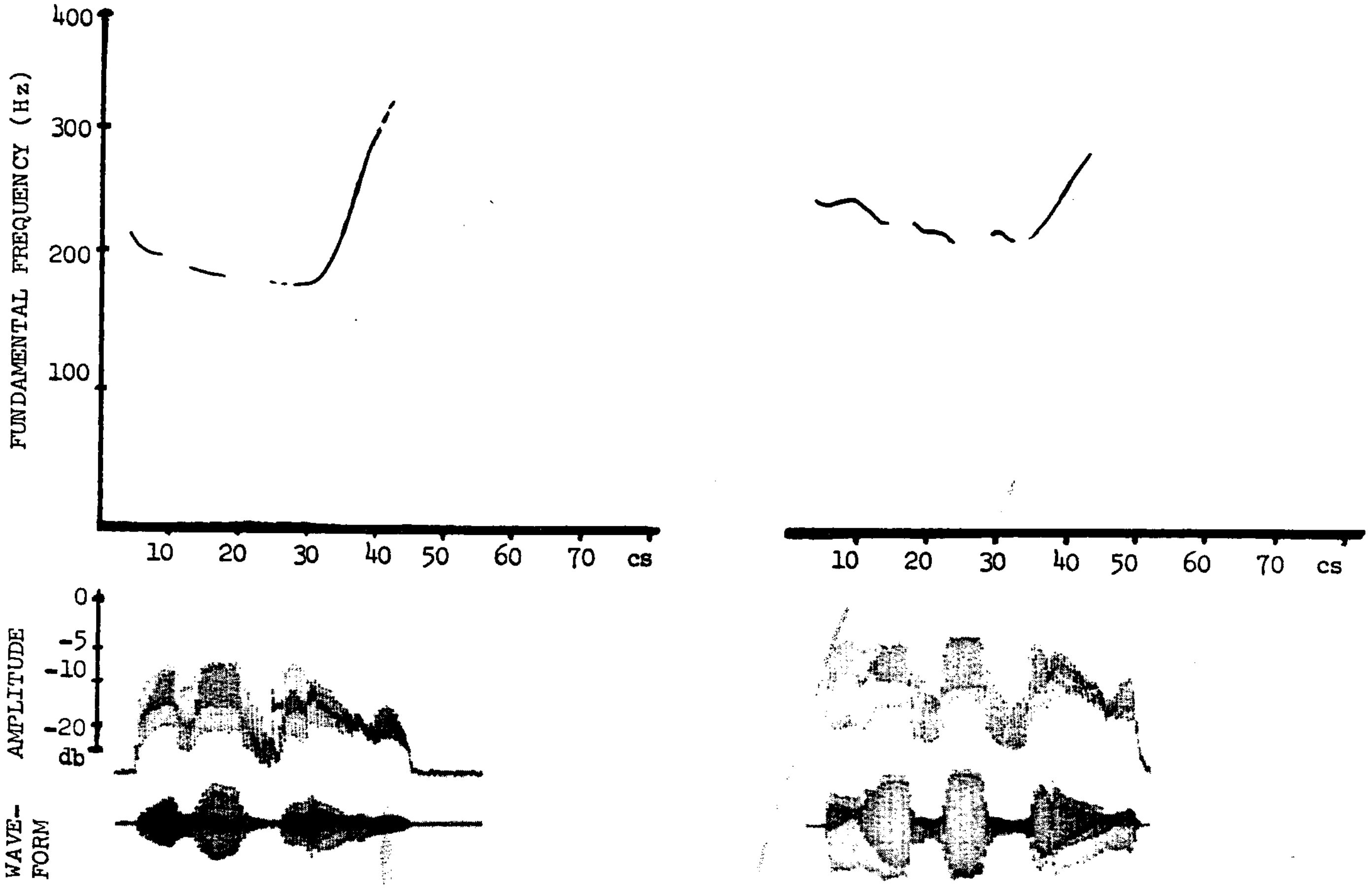


Fig. 4.14 Speaker 8 Were^1

C = 74

Speaker 8 Were^2

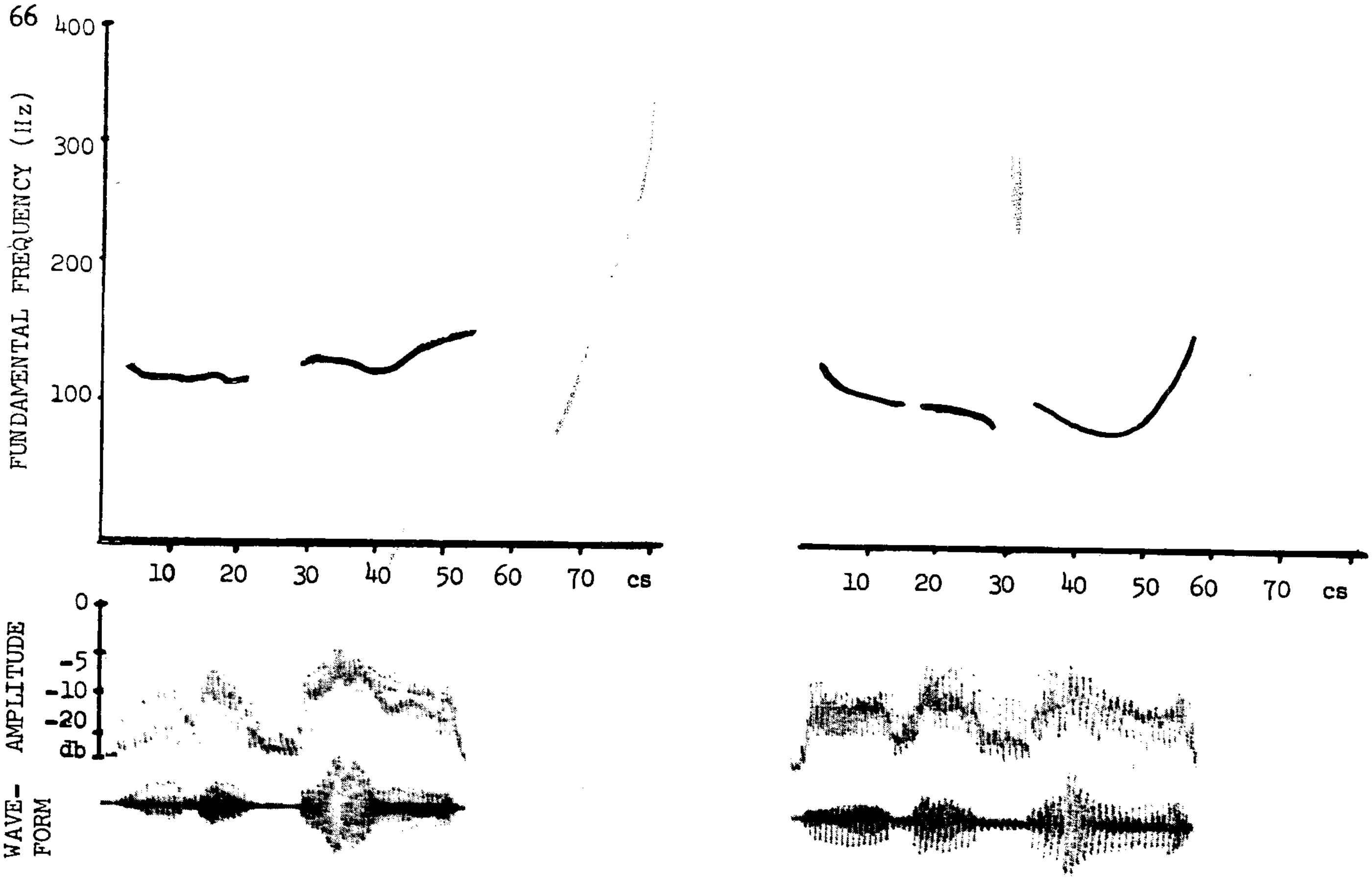


Fig. 4.15 Speaker 1 Were^{1} C = -10 Speaker 1 Were^{2}

Median C-score: 50

Number of Speakers for whom there were

21 or more correct judgments: 10/12

24 or more correct judgments: 9/12

28 or more correct judgments: 9/12

Although this contrast might be said to combine the conceptual and perceptual problems of the previous two contrasts, communication was rather successful here. Speaker 5 (Fig. 4.16) had a contrast between good examples of both *Were*^{2} and *Were*^{3}. Speaker 8's slightly less impressive performance may have been due to a too-sharp rise on *Were*^{2} and a too-sharp fall on *Were*^{3}, but this is only speculation, and we must remember that 41/50 listeners heard her contrast correctly.

We noted earlier (p.51) that test items with marked minus C-scores were suspected of being anomalous. We also observed (p. 63) that Speaker 3 lacked a contrast between *Were*^{1} and *Were*^{3}, or, to be more exact, did not know how to indicate *Were*^{3}. Speaker 3 did have a

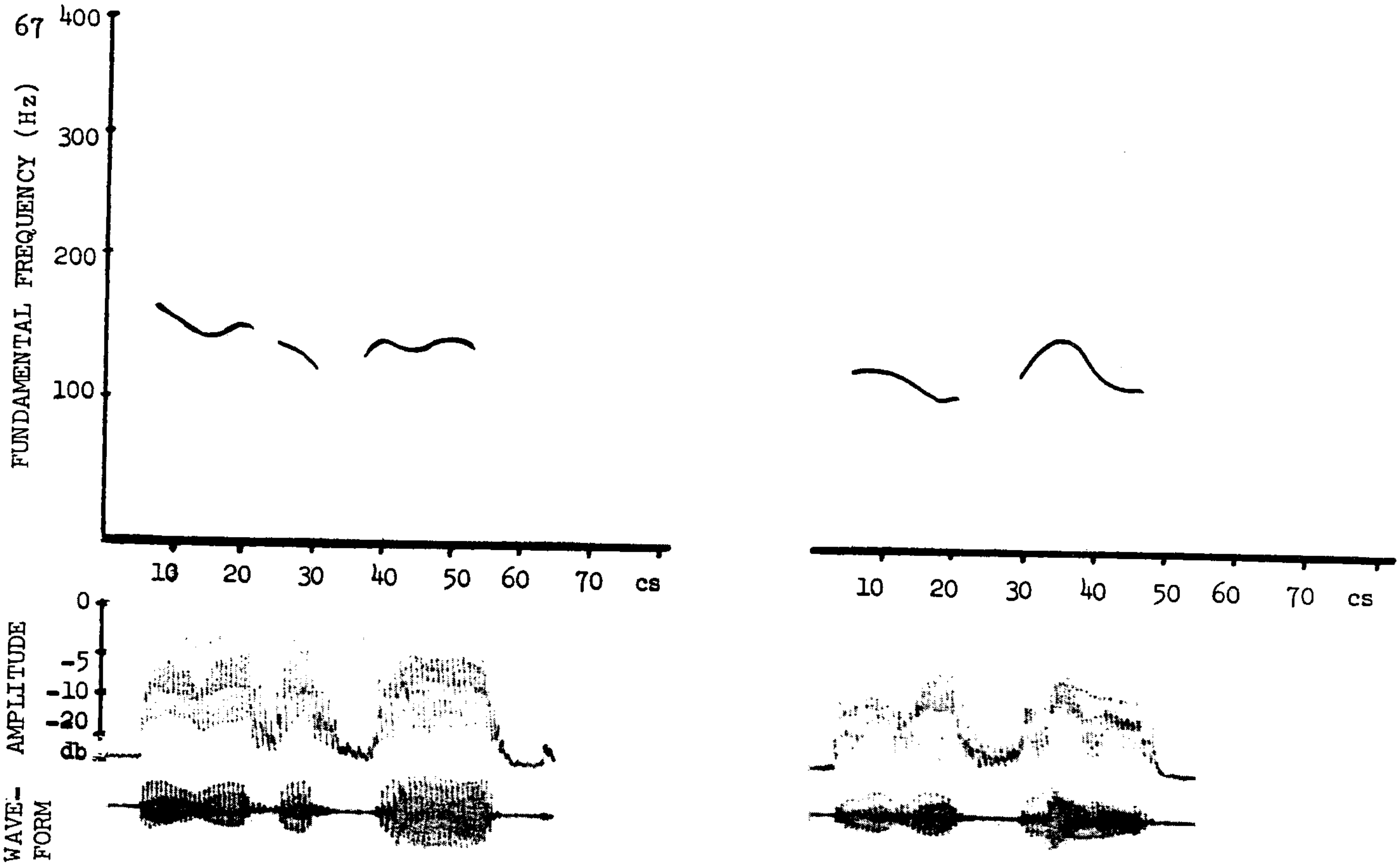


Fig. 4.16 Speaker 5 Were^2 C = 82 Speaker 5 Were^3

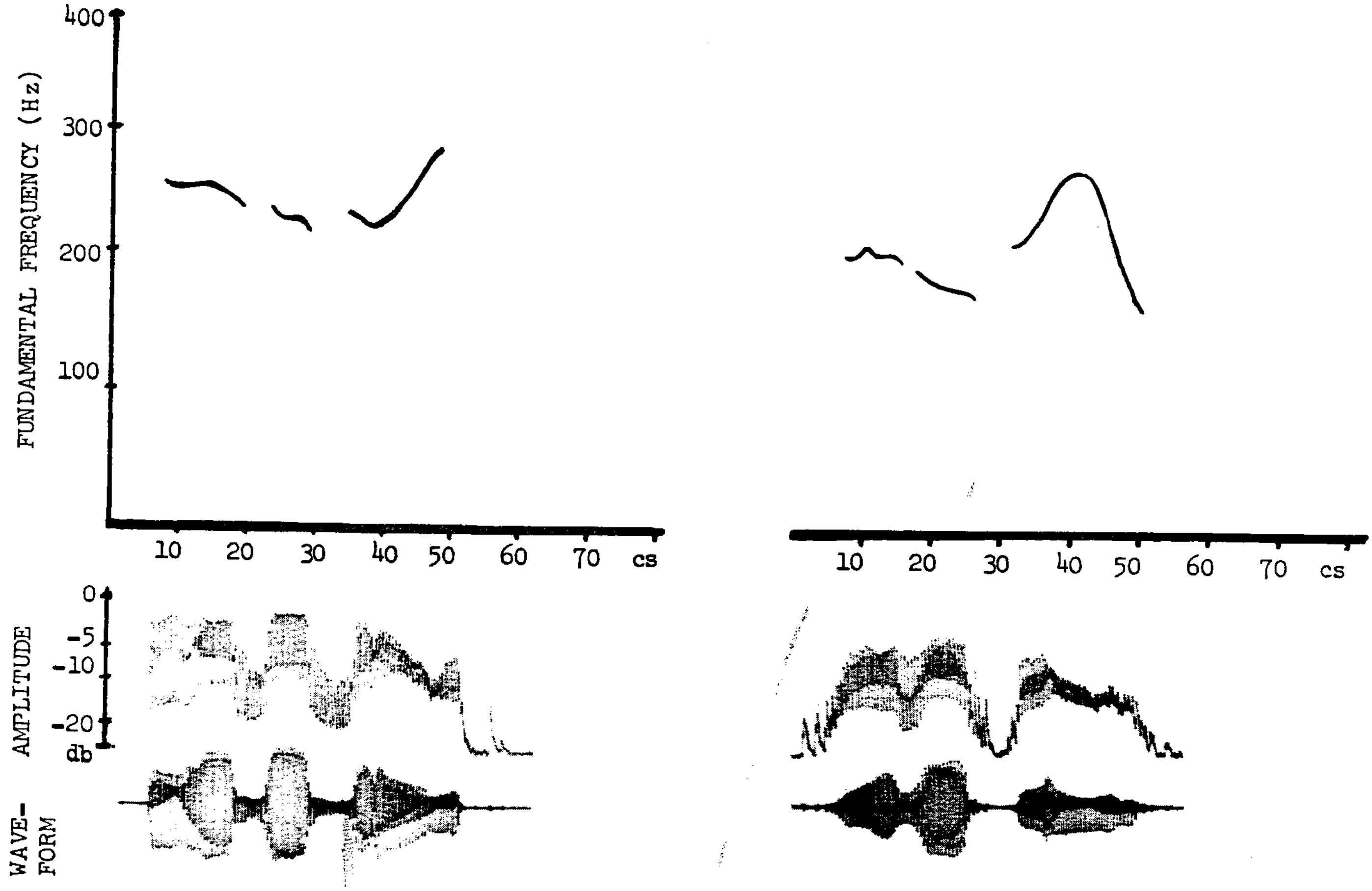


Fig. 4.17 Speaker 8 Were^2 C = 72 Speaker 8 Were^3

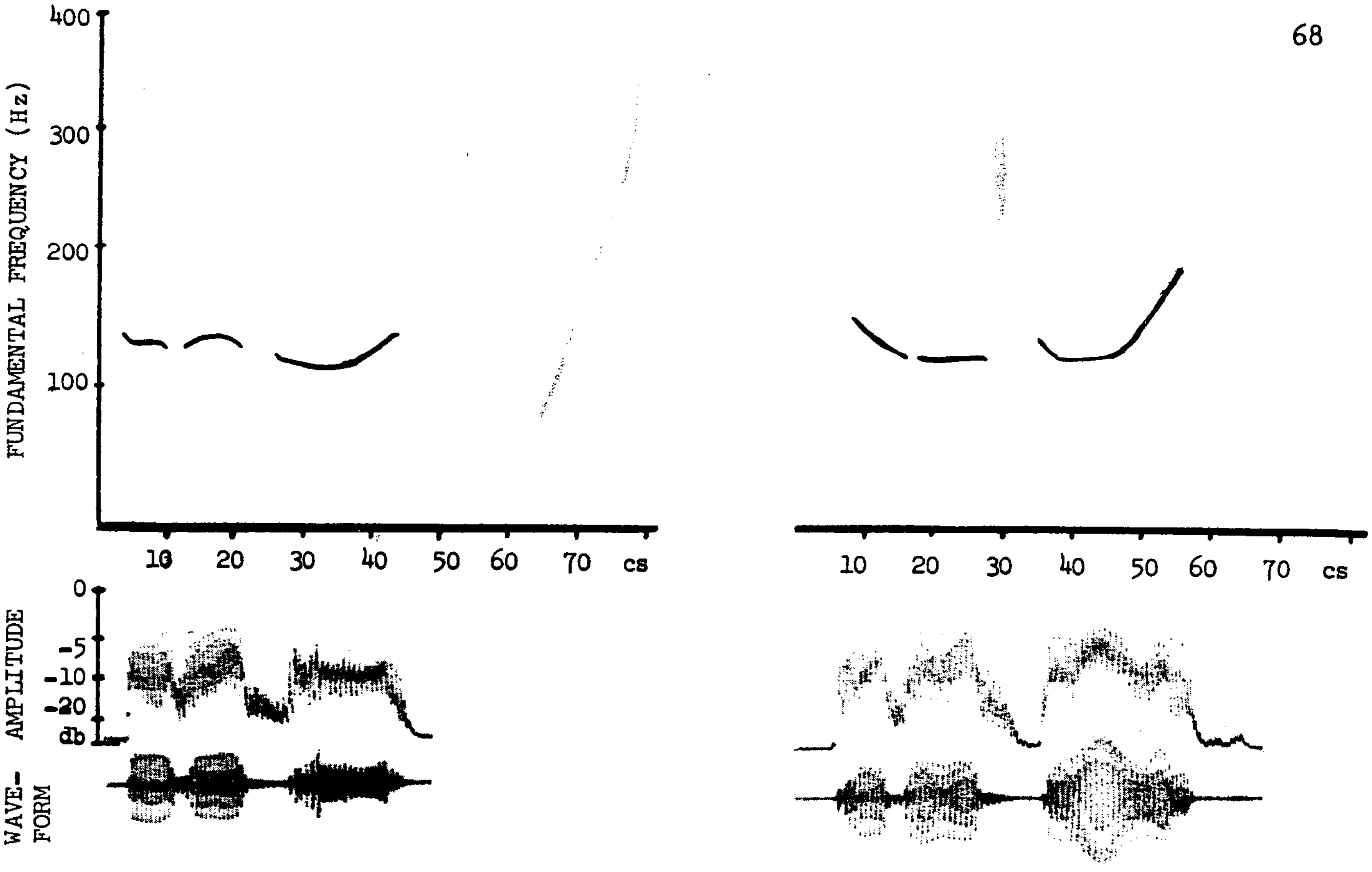


Fig. 4.18 Speaker 3 Were^2 C = -28 Speaker 3 Were^3

good *Were*^{1}/*Were*^{2} contrast, but in the present case the listeners, when confronted with a good *Were*^{2} sample, and a *Were*^{3} which resembled a typical *Were*^{1}, behaved in a manner typical of anomalous stimuli. Of the 50 listeners, 26 labelled both stimuli as *Were*^{2}, while 18, noting a difference in the contours but not certain how to handle it, categorized the stimuli the wrong way around, yielding a C-score of -28.

Yes$^{\text{unemphatic}}$/*Yes*$^{\text{emphatic}}$

NB: With this contrast we begin the analysis of those contrasts elicited by means of cue cards (see pp. 36-39 above).

Mean C-score: 92.83

Median C-score: 94

Number of speakers for whom there were

21 or more correct judgments: 12/12

28 or more correct judgments: 12/12

The contrast between unemphatic and emphatic contours was extremely easy for all speakers. There were four perfect C-scores. Speaker 6 had a contrast roughly similar to that for Speaker 12 (Fig. 4.19), while Speaker 8 achieved her perfect score with a contrast similar to that of Speaker 3 (Fig. 4.20). Lack of emphasis was indicated by contours of moderate length, reduced amplitude, and constricted frequency range. Some of the better *Yes*$^{\text{unemphatic}}$ samples had a very slight nuclear scoop, but none had a marked terminal rise or fall. Speaker 7's "worst" (but still quite successful) performance was probably due to the marked terminal fall and relatively high amplitude of his *Yes*$^{\text{unemphatic}}$ stimulus, which caused nine listeners to place both stimuli in the "emphatic" category. "Emphasis" was indicated by increases in frequency range, amplitude, and duration. All of the better "emphatic" utterances had considerably greater amplitude than the contrasting "unemphatic" stimuli, but there was some evidence of the "trading relationship" noted earlier by Lieberman (1960) and others between frequency range and duration. Thus, Speaker 3, utilizing an extremely high (for a male speaker) nuclear peak on his "emphatic" utterance, needed only a modest increase in duration to achieve a perfect score. On the other hand, Speaker 12, with an "emphatic" contour twice as long as her "unemphatic" utterance, needed only a moderately high nuclear peak to achieve a C-score of 100. There was an additional cue in this case, since Speaker 12 stretched her enunciation of the /y/ in *Yes*$^{\text{emphatic}}$, producing an utterance which might be transcribed as /i^{y}yɛ́:s/. This enunciation might be likened to the heightening and/or stretching of the first syllable in the "emphatic" utterance of *ridiculous*, which was done by ten of the speakers.

Ridiculous$^{\text{unemphatic}}$/*Ridiculous*$^{\text{emphatic}}$

Mean C-score: 92.50

Median C-score: 94

Number of speakers for whom there were

21 or more correct judgments: 12/12

28 or more correct judgments: 12/12

The performance of the speakers on this test category almost exactly matched the extremely effective level of communication shown on the *Yes*$^{\text{unemphatic}}$/*Yes*$^{\text{emphatic}}$ category. Surprisingly, amplitude was only a minor cue here, since the speakers apparently could not bring themselves to enunciate q word with the semantic force of *ridiculous* without giving it a greater amount of amplitude than they did on the unemphatic utterance of *yes*. Nor were there great differences in total duration between the unemphatic and emphatic samples of

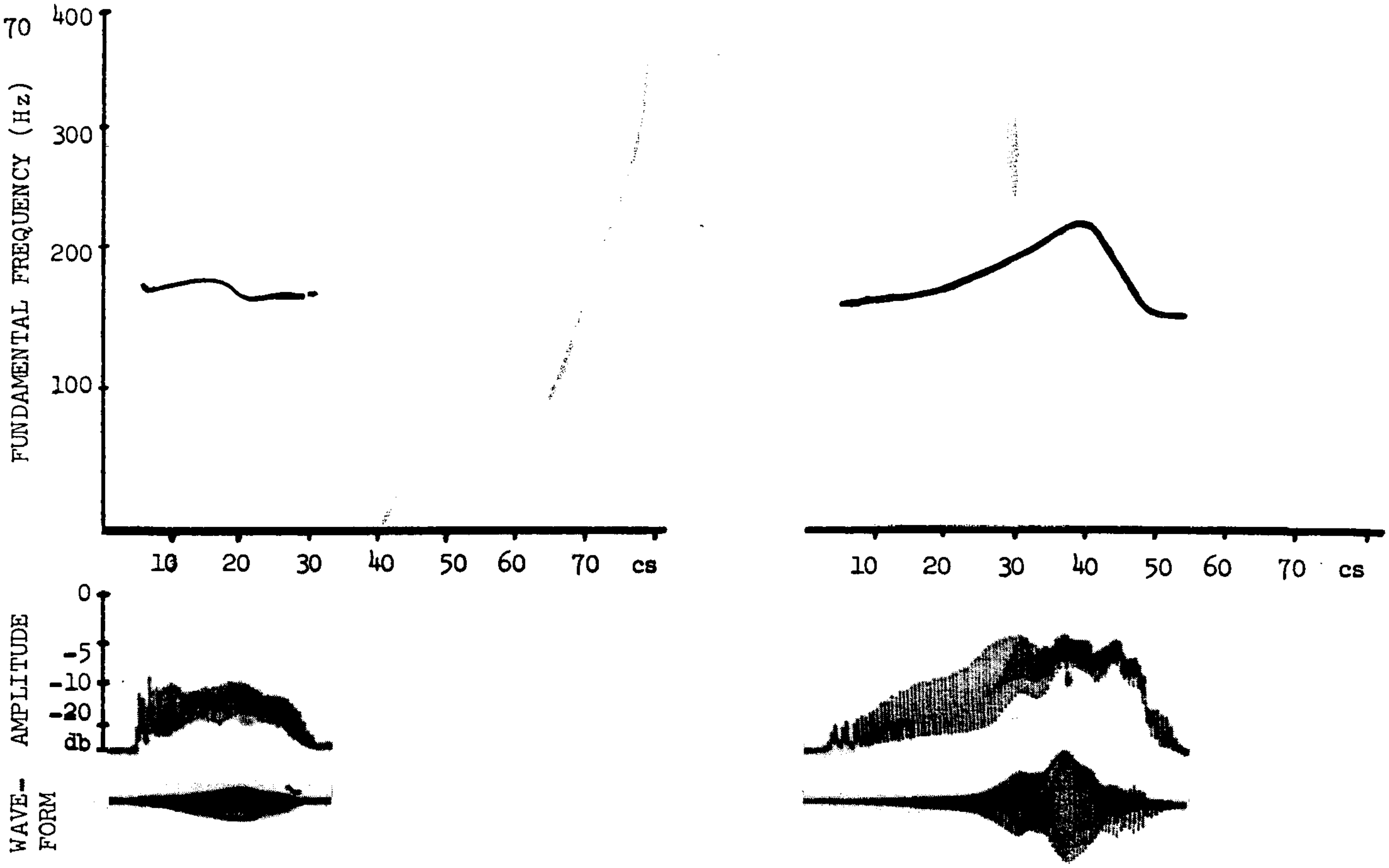


Fig. 4.19 Speaker 12 Yes^{unemphatic} C = 100 Speaker 12 Yes^{emphatic}

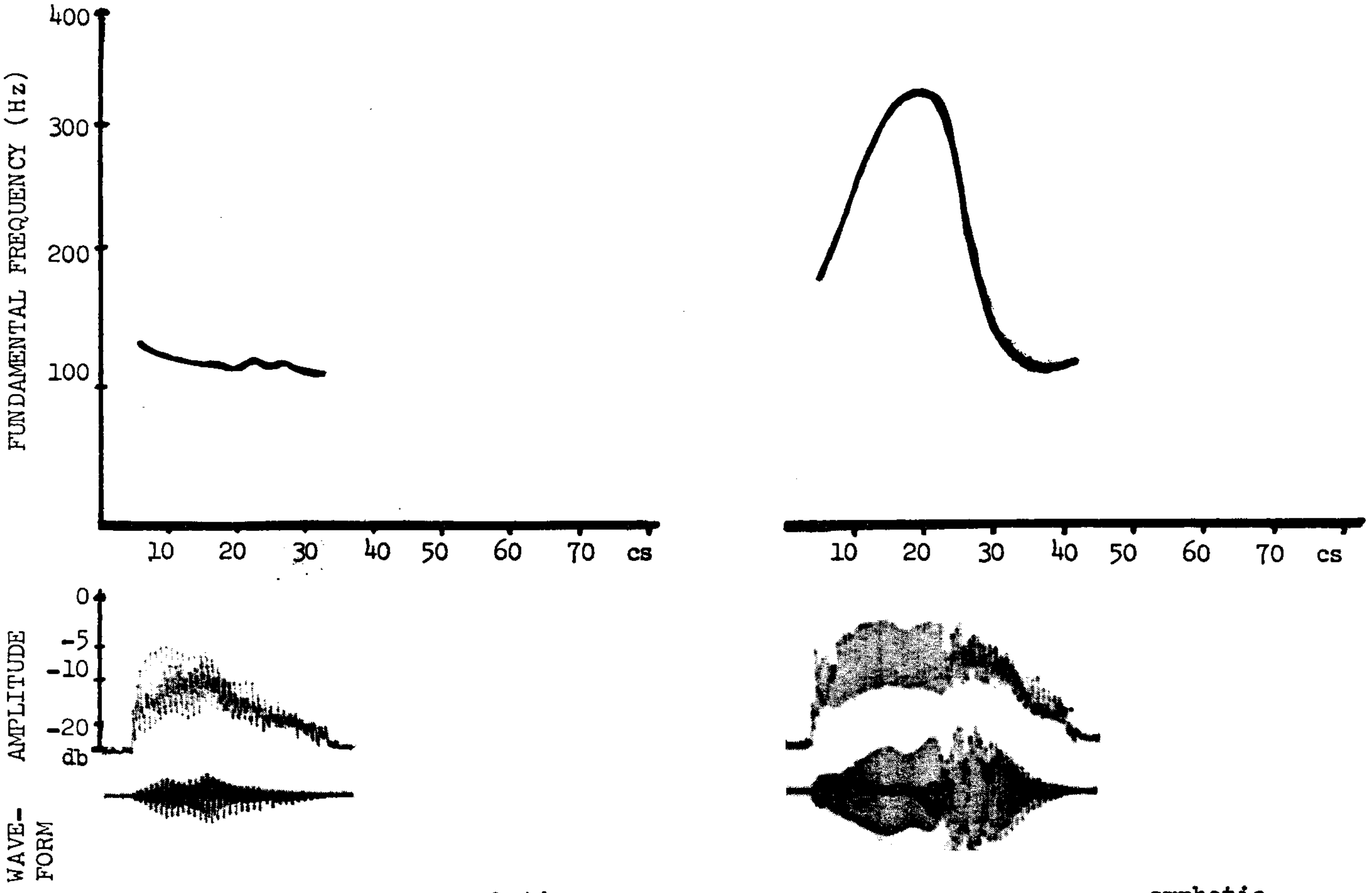


Fig. 4.20 Speaker 3 Yes^{unemphatic} C = 100 Speaker 3 Yes^{emphatic}

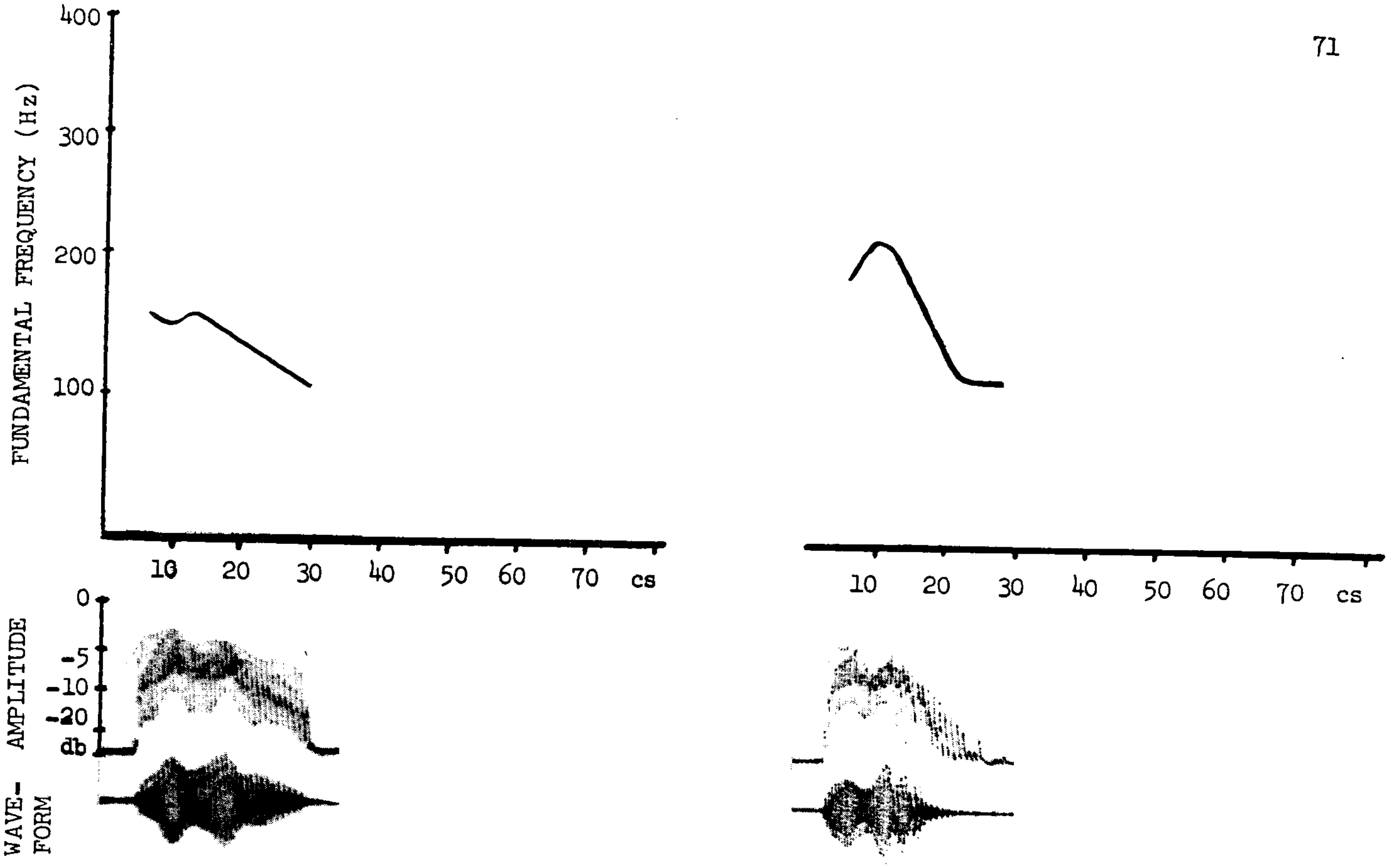


Fig. 4.21 Speaker 7 Yes$^{\text{unemphatic}}$ C = 78 Speaker 7 Yes$^{\text{emphatic}}$

ridiculous. Rather, the emphatic nature of the utterance is indicated by a change in the relationship between the first two syllables. In the unemphatic enunciations, the first syllable is pronounced with a lax vowel, and is closely followed by the second, stressed syllable, the frequency range of which is very similar to that for the first syllable. In the emphatic enunciations, however, there is a much longer juncture between the end of the first syllable and the beginning of the second, nuclear-peaked syllable. In some cases, the first syllable has a vowel of greater duration, and in five of the best examples, this greater duration was accompanied by a change in vowel quality from /i/ (unemphatic) to /iy/ (emphatic). Lastly, the most effective examples of *Ridiculous*$^{\text{emphatic}}$ also showed a considerable heightening of the frequency of the first syllable, so that it rose higher than the second syllable, which still bore a recognizable primary stress. As Figs. 4.22 and 4.23 make clear, Speaker 6 and Speaker 2 both made use of heightened first syllables and longer junctures, and Speaker 6 had a longer first syllable for her emphatic utterance. The illustrations do not show that both also changed the vowel quality from /i/ to /iy/ in the first syllable, a change which, together with the longer juncture, made Speaker 2's first syllable in *Ridiculous*$^{\text{emphatic}}$ seem longer than the first syllable of her unemphatic utterance.

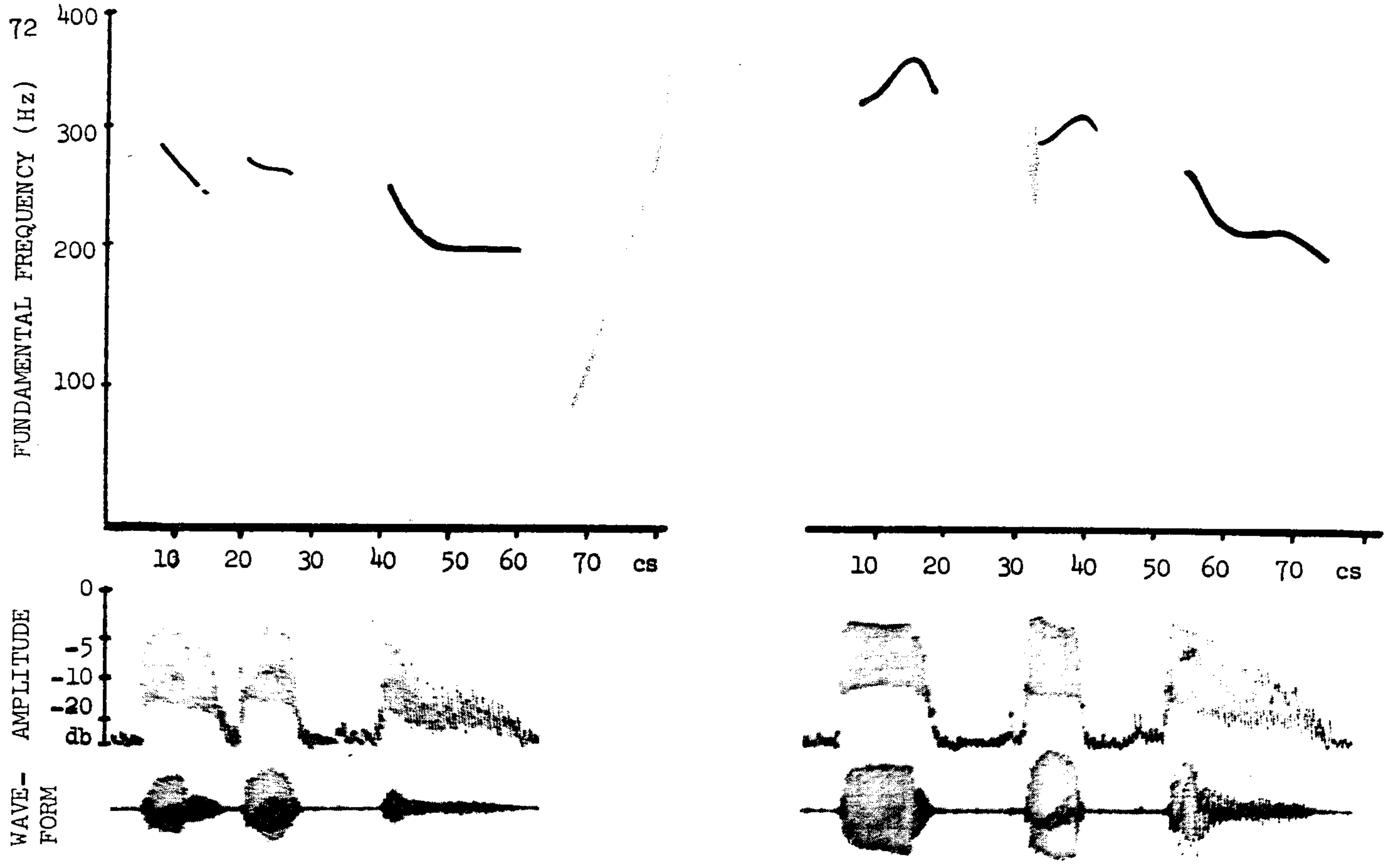


Fig. 4.22 Speaker 6 Ridiculous$^{\text{unemphatic}}$ C = 100 Speaker 6 Ridiculous$^{\text{emphatic}}$

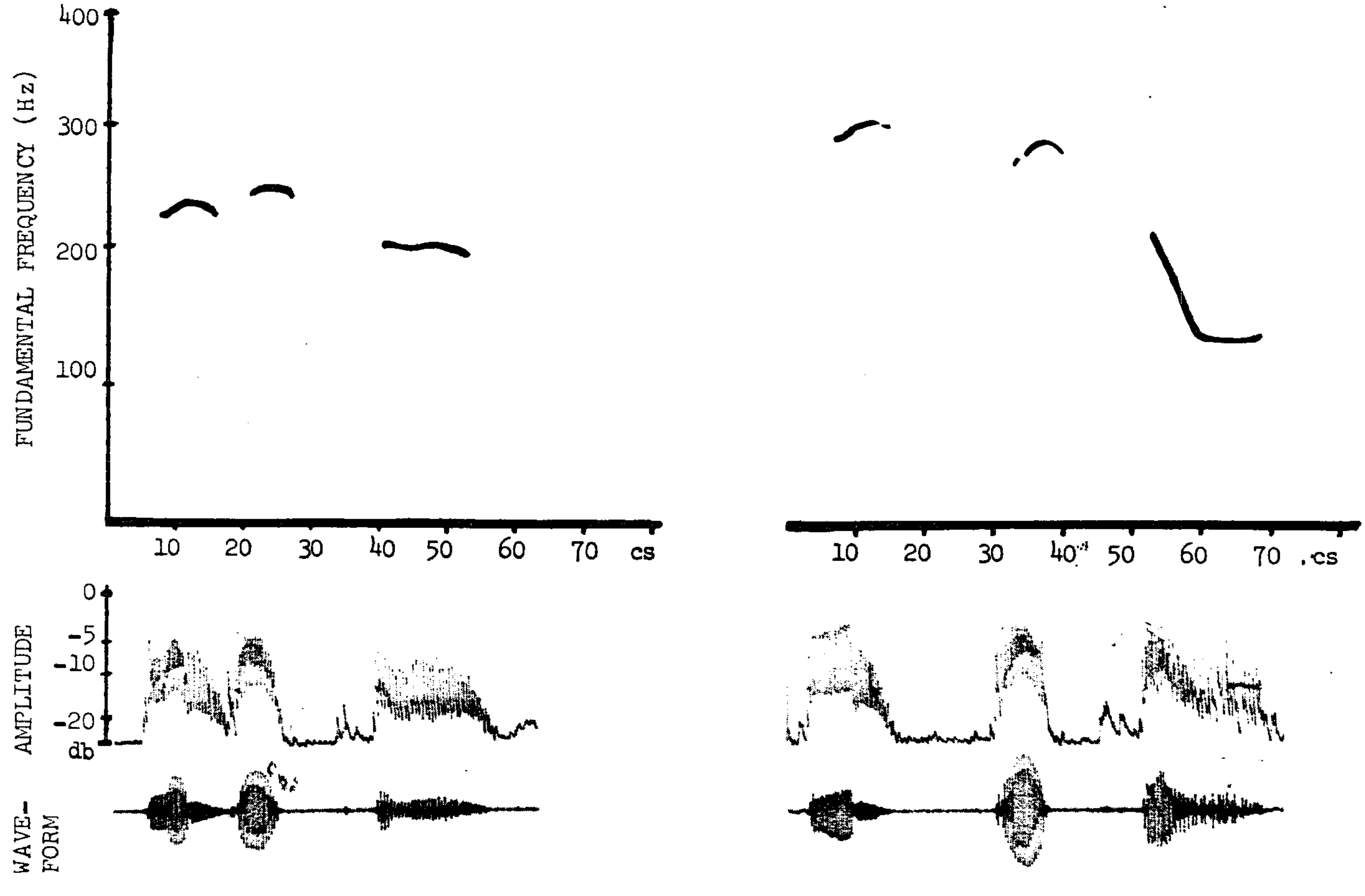


Fig. 4.23 Speaker 2 Ridiculous$^{\text{unemphatic}}$ C = 100 Speaker 2 Ridiculous$^{\text{emphatic}}$

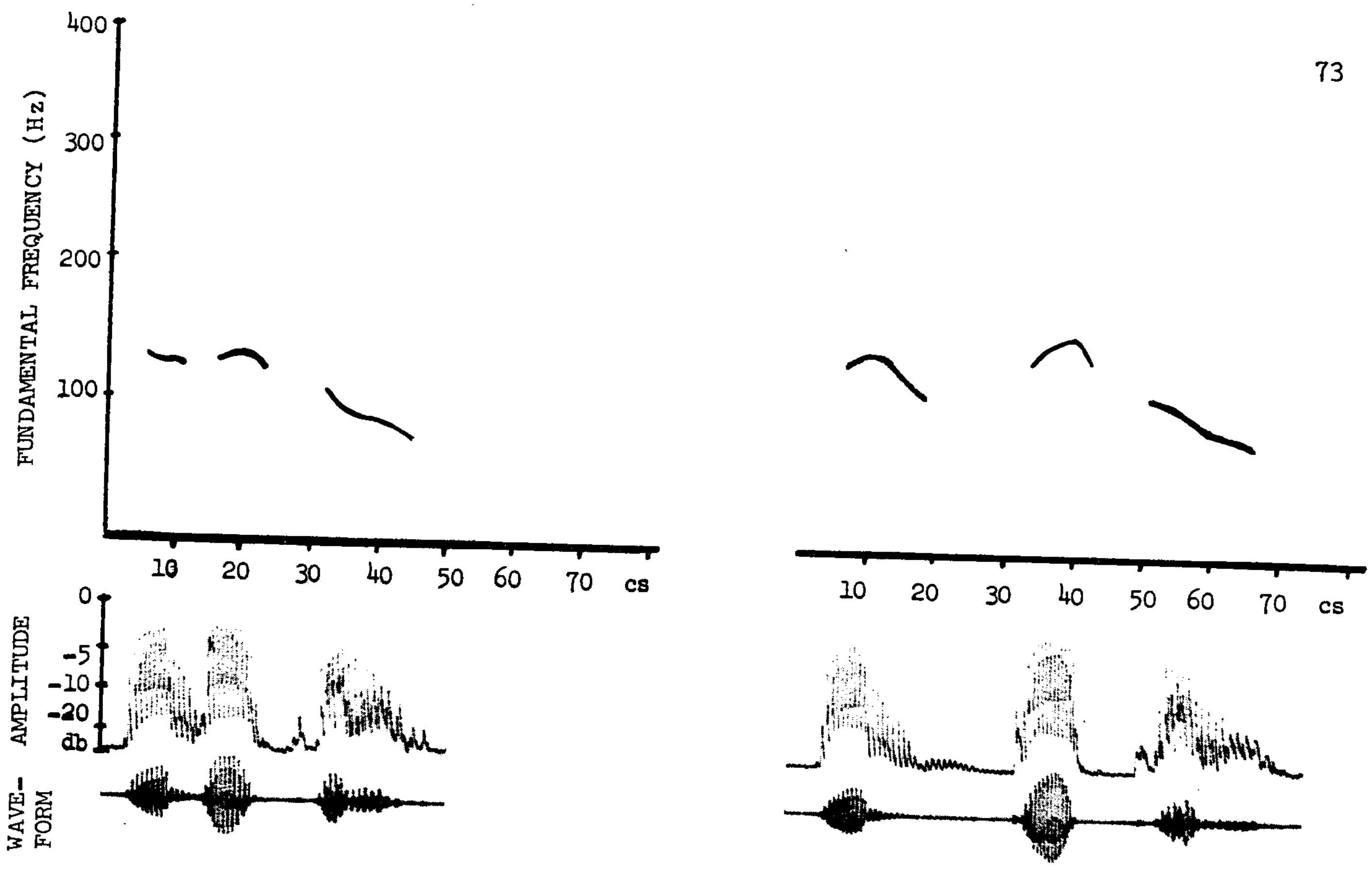


Fig. 4.24 Speaker 9 Ridiculous$^{\text{unemphatic}}$ C = 84 Speaker 9 Ridiculous$^{\text{emphatic}}$

Speaker 9's C-score of 84 is surprisingly high for a "worst" performance. Of the 50 listeners, 43 categorized his stimuli correctly, only one incorrectly. However, six listeners placed both stimuli in the "emphatic" category, perhaps because of the relationship between the low terminal fall and the higher earlier portions of his "unemphatic" utterance, which gave this utterance an air of "finality" and, perhaps by extension, "emphasis".

Summary of Results for Syntactic Test Categories

There were eight syntactic test categories: *Yes*^{1}/*Yes*^{4} (*yes* as a statement versus *yes* as a "repetitive" question), *Yes*^{1}/*Yes*^{2} (statement/incomplete statement), *Yes*^{3}/*Yes*^{4}(a continuative "telephone" *yes*/"repetitive" question), *Were*^{1}/*Were*^{3} (*were they black* as "normal" question and as subjunctive, meaning " they were black..."), *Were*^{1}/*Were*^{2} ("normal" question versus the first part of an incomplete, *or*-type question), *Were*^{2}/*Were*^{3}, *Yes*$^{\text{unemphatic}}$/*Yes*$^{\text{emphatic}}$, and *Ridiculous*$^{\text{unemphatic}}$/*Ridiculous*$^{\text{emphatic}}$. The last two categories used items elicited by means of cue cards (see pp. 36-39 above). The test items for the first six categories were excised from a dialogue (see pp. 32-36).

The most difficult category was *Yes*^{3}/*Yes*^{4}, on which only six of the twelve speakers were able to communicate at a significant level. All of the other categories showed very effective communication, with the "unemphatic"/"emphatic" category being the easiest. The most surprising results stemmed from the fact that the great majority of the speakers were able to achieve effective communication in the theoretically "difficult" *Were*^{1}/*Were*^{2}/*Were*^{3} categories.

The data from these eight categories, and from the "emotional" categories discussed in the next chapter, show that even in the absence of other context, most speakers are able to communicate a great deal of information by intonation alone. They also suggest that some speakers are more effective at this communication than others. The implications of these findings will be discussed in the last chapter of this study.

Chapter 5: Emotional Aspects of American English Intonation

The extremely effective performance of the speakers in the "emotional" test categories came as an agreeable surprise. We had anticipated that the "calm"/"angry" distinction would be very easy for all the speakers to communicate, and included the "anger"/"contained anger" test as a theoretically "difficult" category. But even on this category, ten of the twelve speakers were able to communicate at an extremely significant level. Two very anomalous test items (Speaker 2's *Yes*$^{\text{bored}}$/*Yes*$^{\text{interested}}$ and Speaker 10's *Yes*$^{\text{agreeable}}$/*Yes*$^{\text{disagreeable}}$) complicated the results slightly, but even including the negative communication of those items, the results were quite impressive, particularly in view of the fact that the speakers had to communicate these emotional attitudes entirely through samples of the word *yes*, with no additional communicative context, except for the "reference tone" supplied by the phrase *Today is Monday* (see p. 44 above).

In the communication of these emotional attitudes, paralinguistic cues (usually manifested as alterations in voice quality) frequently played an important rôle. Because of our interest in eliciting speech samples which would be as natural as possible, we did not make use of devices such as air flow meters, or techniques such as cineradiology or electromyography which might have given reliable evidence concerning the means of achieving these alterations in voice quality. We have therefore had to rely primarily upon acoustic impressions, and upon a subjective vocabulary (e.g., "hard edge to the voice", "tight, squeezed voice quality", etc.) for communicating these impressions to the reader. We regret the necessity of using such "imitation labels" (Pike, 1945). In a future study using more controlled stimuli, we hope to remedy this limitation.

Discussion of Emotional Test Categories

Yes$^{\text{belief}}$/*Yes*$^{\text{disbelief}}$

Mean C-score: 66.83

Median C-score: 66

Number of speakers for whom there were

21 or more correct judgments: 12/12

28 or more correct judgments: 12/12

The stimuli produced by the speakers for this category are characterized by a great deal of uniformity among the "belief" samples, and a wide diversity of approaches to communication on the "disbelief" utterances. All of the speakers produced a basic "statement" type of contour for *Yes*$^{\text{belief}}$, and all of the more successful contrasts contained *Yes*$^{\text{belief}}$ contours with gently peaked nuclei and sustained, slightly rising, or slightly falling terminal segments, such as are seen in Figs. 5.1-2. However, the "disbelief" utterances display a variety not only in fundamental frequency contours, but also in paralinguistic means for communicating disbelief. The most successful communicators of this contrast were those who combined an appropriate frequency contour with effective paralinguistic cues.

The most effective basic contour for signalling disbelief was that used by Speaker 4 (Fig. 5.1): a high, level contour with a sharp upturn at the very end. Judging from the slightly less successful versions of the same basic contour used by Speakers 2 and 6, it would appear that all three factors (height, level nuclear portion, and sharp terminal upturn) are important. Speaker 4's performance was augmented by the effective use of a tight, somewhat wobbling tone (probably due to pharyngeal tightening) with her "disbelief" stimulus.

The other successful maneuver for communicating disbelief consisted of using a more neutral f_o contour, but drawling it out, and using the greater length of the "disbelief" contour as a vehicle for communicating various paralinguistic cues for disbelief. Thus Speaker 12 (Fig. 5.2) deliberately lengthened her "disbelief" contour, and added to it a tight, vibrato-like tone which was apparently produced by a relatively high variety of creaky voice. The weakest performers on this contrast all failed because they utilized f_o contours for "disbelief" which were too much like ordinary statement or question contours, and added paralinguistic cues which were too subtle to be clearly perceived. Speaker 11's "disbelief" contour (Fig. 5.3) is not sufficiently level across the nuclear portion to sound unlike a normal statement, nor is this contour different in length from his "belief" utterance (in general, listeners equated brevity with "belief", long, drawled-out contours with "disbelief"). Lastly, his attempt to put a "biting" tone on his "disbelief" utterance was much too subtle. Almost all of the attempts at differentiating the two contours by means of a hard biting tone on *Yes*$^{\text{disbelief}}$ failed, perhaps because the speakers did not avoid tensing their vocal tract muscles in advance, and thereby producing such a tone on *both* the "belief" and "disbelief" stimuli. Paralinguistic cues which were generally effective for communicating disbelief were the high, wobbling tone previously mentioned (Speaker 4), which reached falsetto-

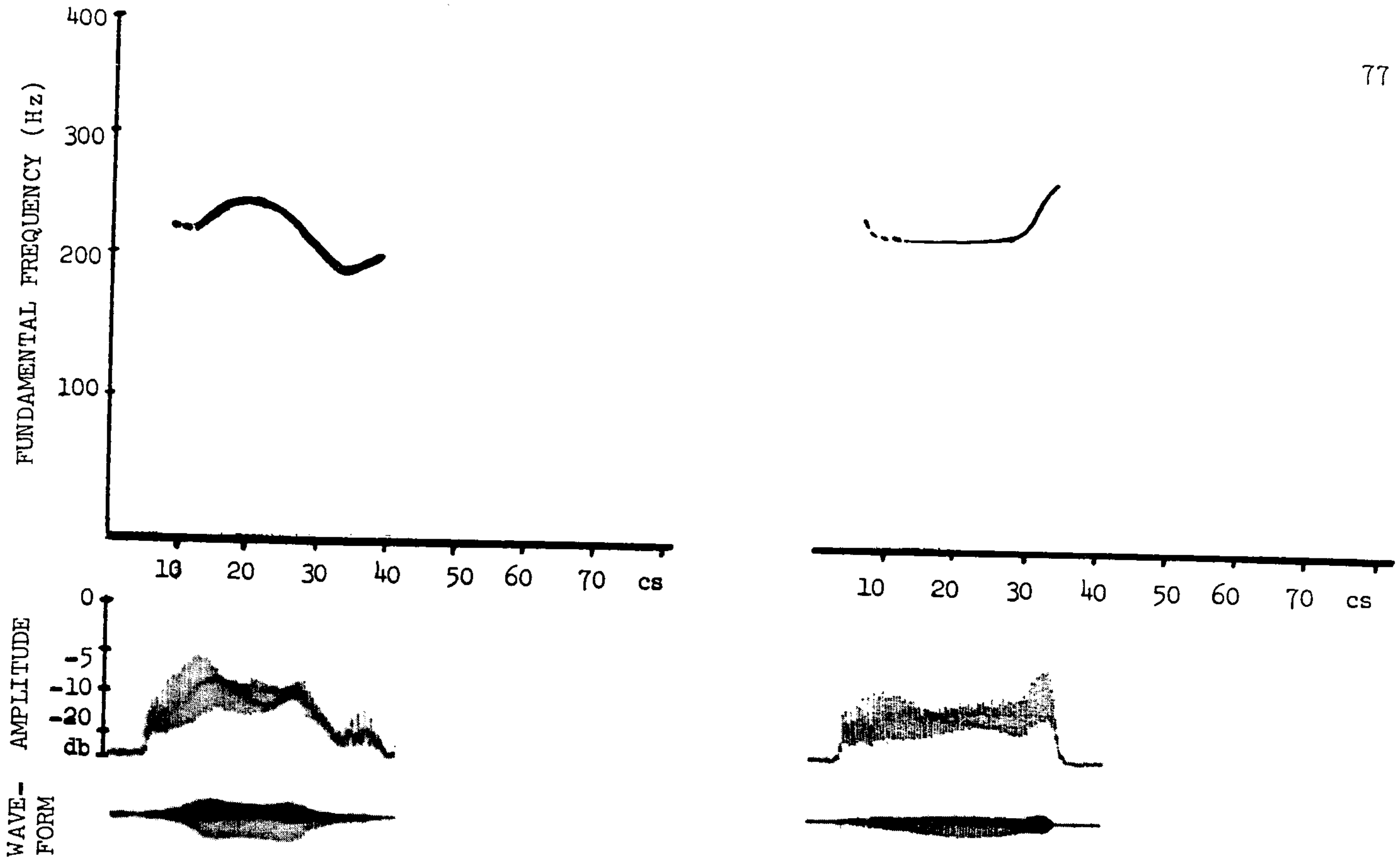


Fig. 5.1 Speaker 4 Yes^{belief}

C = 94 Speaker 4 Yes^{disbelief}

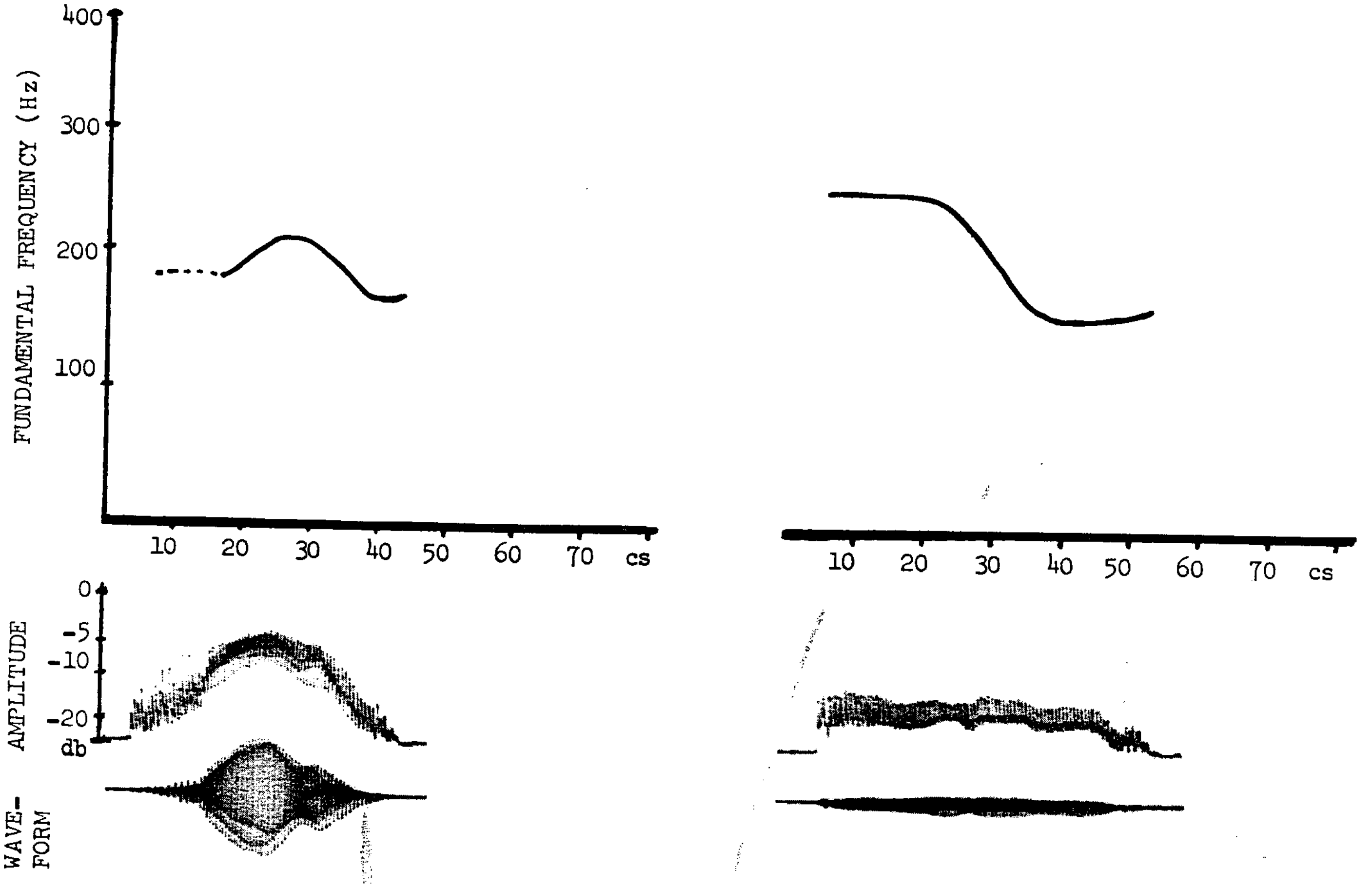


Fig. 5.2 Speaker 12 Yes^{belief}

C = 92 Speaker 12 Yes^{disbelief}

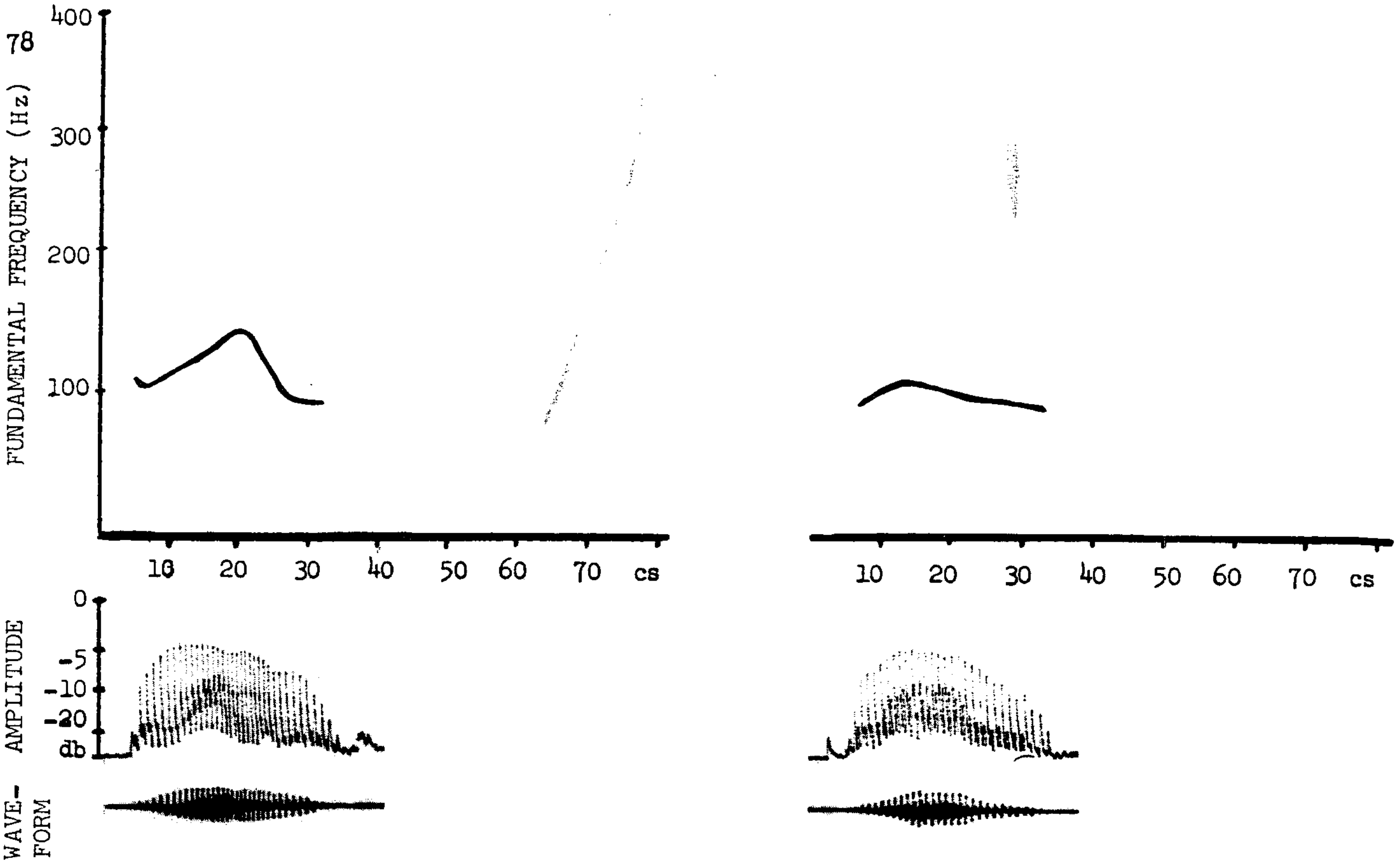


Fig. 5.3 Speaker 11 $\text{Yes}^{\text{belief}}$ C = 40 Speaker 11 $\text{Yes}^{\text{disbelief}}$

like dimensions in the case of Speaker 7 (C=90), and a drawling-out of the initial /y/ in *yes* (as in the example on p. 71 above), manifested by Speaker 1 (C=86) and Speaker 2 (C=80). We believe that the delicate interaction of f_o contour and paralinguistic cues illustrated above represents an extraordinarily interesting and challenging area of research for synthetic speech experimentation in the approaching decade.

$\textit{Yes}^{\text{bored}}/\textit{Yes}^{\text{interested}}$

Mean C-score: 81.50

Mean C-score for top 11 speakers: 94.55
(Speaker 2 eliminated)

Median C-score: 96

Number of speakers for whom there were

21 or more correct judgments: 11/12

28 or more correct judgments: 11/12

Because of the extremely anomalous stimuli produced by Speaker 2, we have supplied for this contrast not only the two "best" test items (Figs. 5.4-5) and the one "worst" (Speaker 2, Fig. 5.7) but also a "typical worst" item (Fig. 5.6), so that the reader might obtain a better idea of the nature of this contrast.

Examining the eleven non-anomalous test items, we find that communication of this contrast was at an extremely high level. Furthermore, although all of the speakers attempted to produce a difference in voice quality as an additional means of differentiating the two stimuli, most of these paralinguistic cues were subtle to the point of being barely audible, and the major amount of communication seemed to be carried by the f_o contour itself. As Figs. 5.4-5 show, a very lively contour best signals "interest", while the "bored" stimulus ideally has a very narrow-ranged contour, dropping slightly from a level precontour to a level nuclear portion, followed by a sustained or very slightly rising terminal segment. Additionally, the "bored" stimulus typically shows a marked decrease in amplitude toward the end of the contour, while the "interested" utterance usually has a definite increase in amplitude over the terminal section.

Concerning the paralinguistic cues, the speakers typically attempted to communicate (during the "bored" utterance) the acoustic flavor of an incipient or stifled yawn. Some did this with a slight vibrato effect. Others used a slightly "breathy" voice quality. The most successful was Speaker 12, who augmented her distinctive contours (Fig. 5.5) with a breathy enunciation of *Yes*$^{\text{bored}}$, followed immediately by a forceful exhalation, further communicating a general air of boredom and, possibly, disgust. However, because all of the non-anomalous test items were so successful in communicating the "bored"/"interested" contrast, and because we have no examples of stimuli which differ only according to a presence or absence or paralinguistic cues for boredom, we cannot say how effective these cues were, nor even whether they were at all necessary. Here again, the manipulation of parameters possible in synthetic speech experiments would appear to offer exciting possibilities.

Speaker 2's stimuli (Fig. 5.7) are anomalous on at least two counts. First, her "bored" contour sounds a bit too "interested". Far more serious, however, is the very strange contour she produced as an example of *Yes*$^{\text{interested}}$. The constricted f_o range, the high amplitude peaking at contour end, and the break in voice all indicate an unpleasant sort of agitation far more typical of a category such as "disagreeableness" or "contained anger". Confronted with this pair of stimuli, 31 of the lsiteners categorized them the wrong way around, while the other 19 listeners labelled both stimuli as examples of *Yes*$^{\text{bored}}$, yielding a C-score of -62.

The utterances produced by our "typical worst" speaker (Speaker 4, Fig. 5.6) were correctly categorized by 40 listeners, with only one

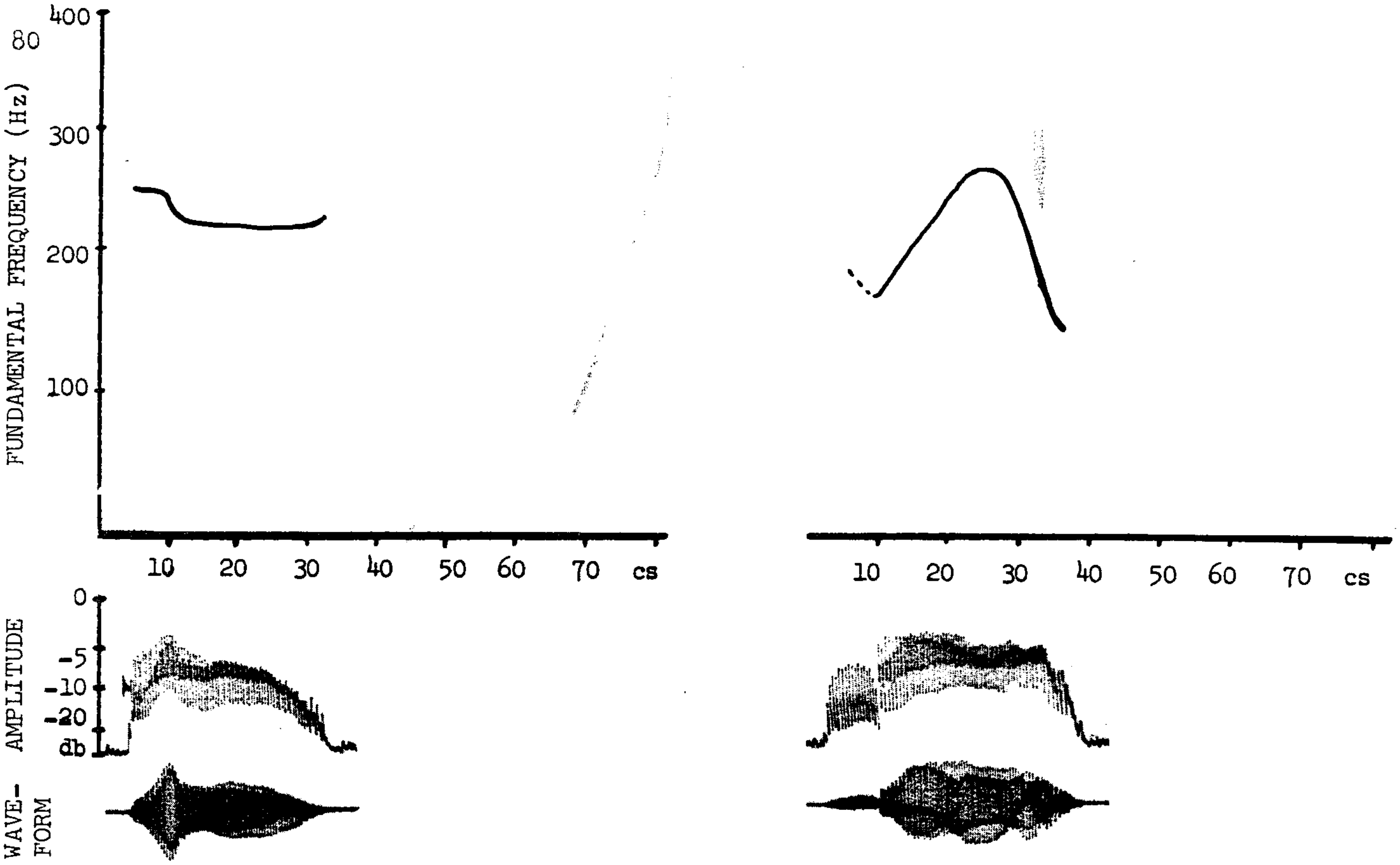


Fig. 5.4 Speaker 8 Yes^{bored} C = 100 Speaker 8 Yes^{interested}

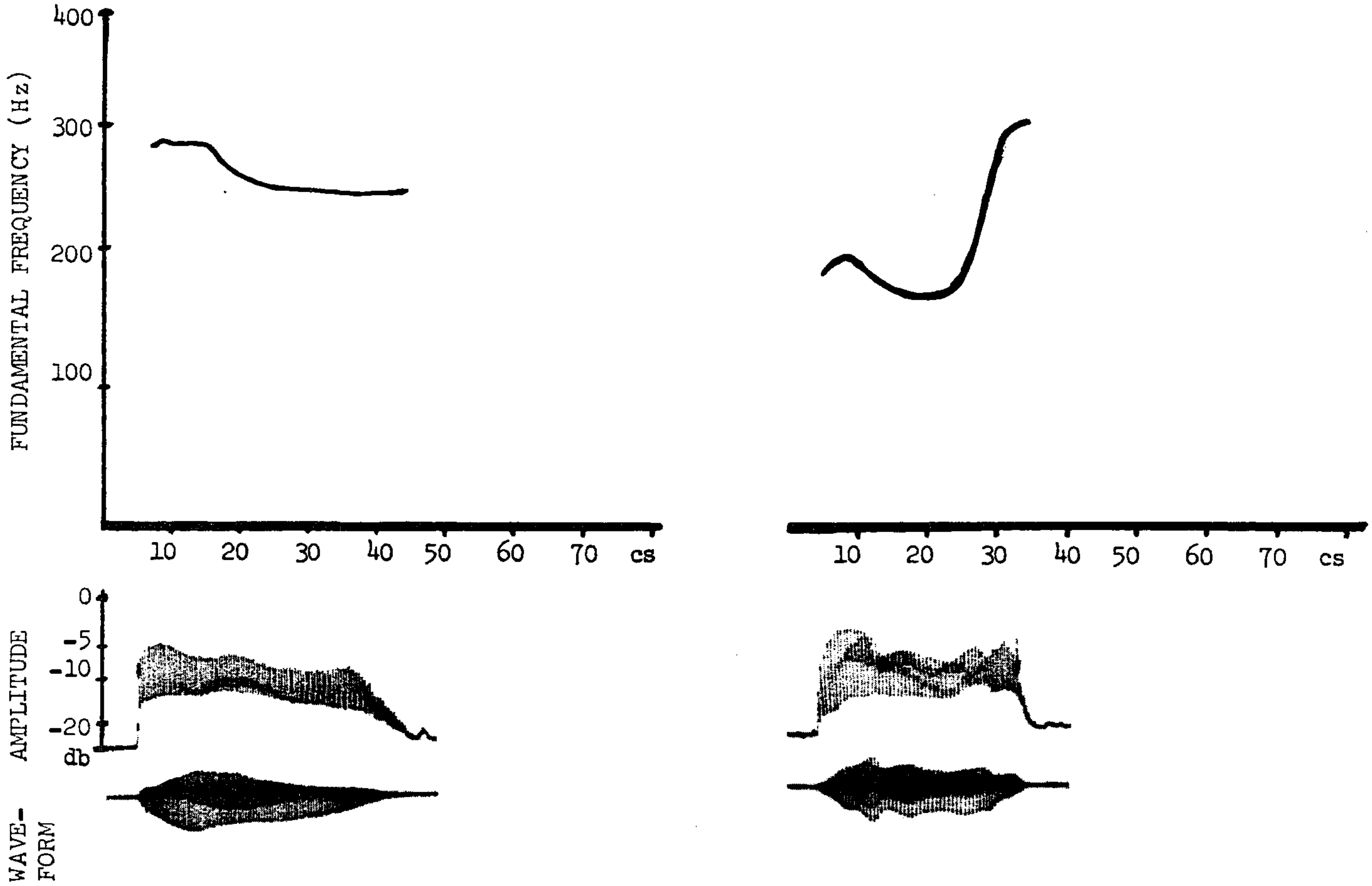


Fig. 5.5 Speaker 12 Yes^{bored} C = 100 Speaker 12 Yes^{interested}

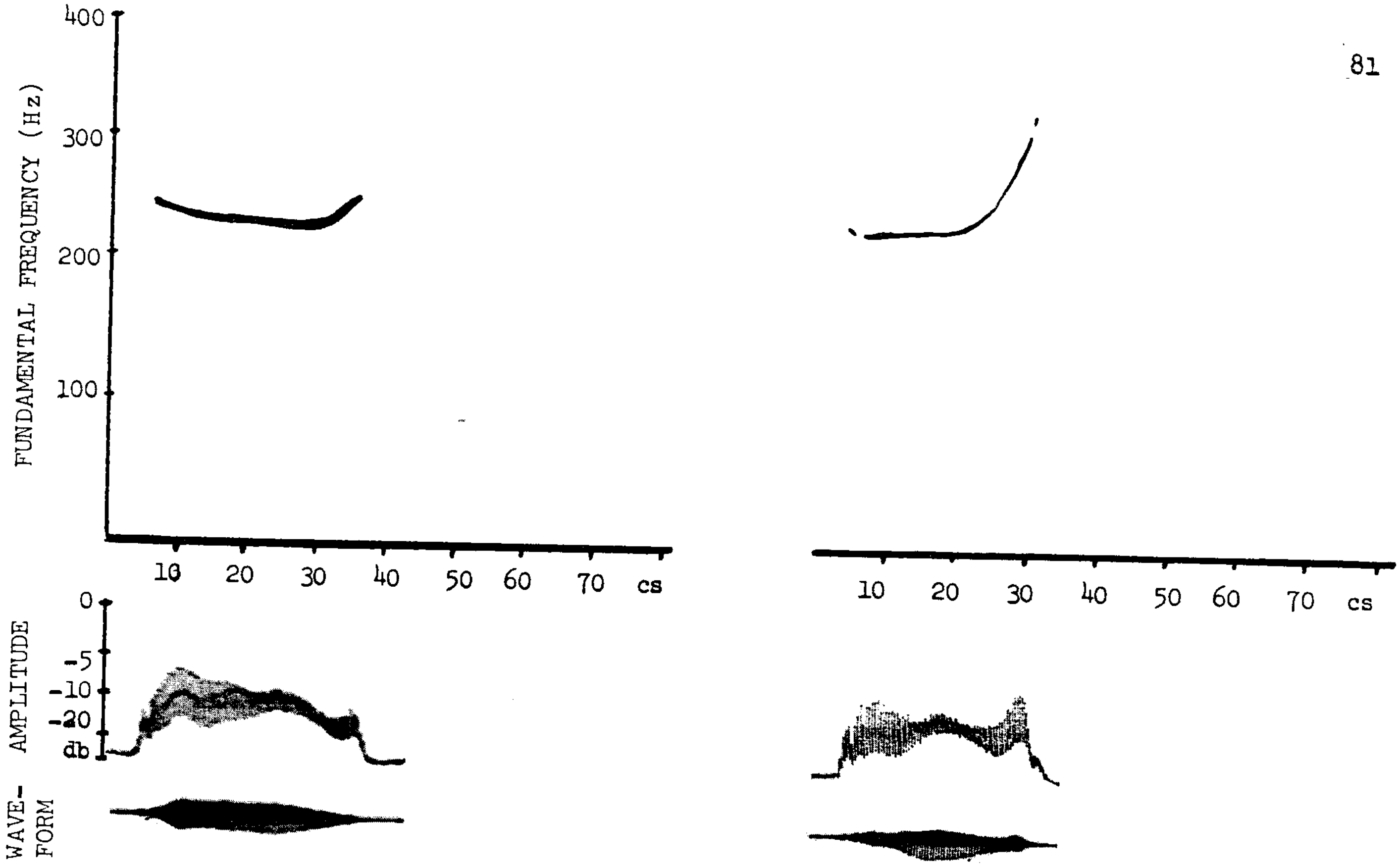


Fig. 5.6 Speaker 4 $\text{Yes}^{\text{bored}}$

C = 78 Speaker 4 $\text{Yes}^{\text{interested}}$

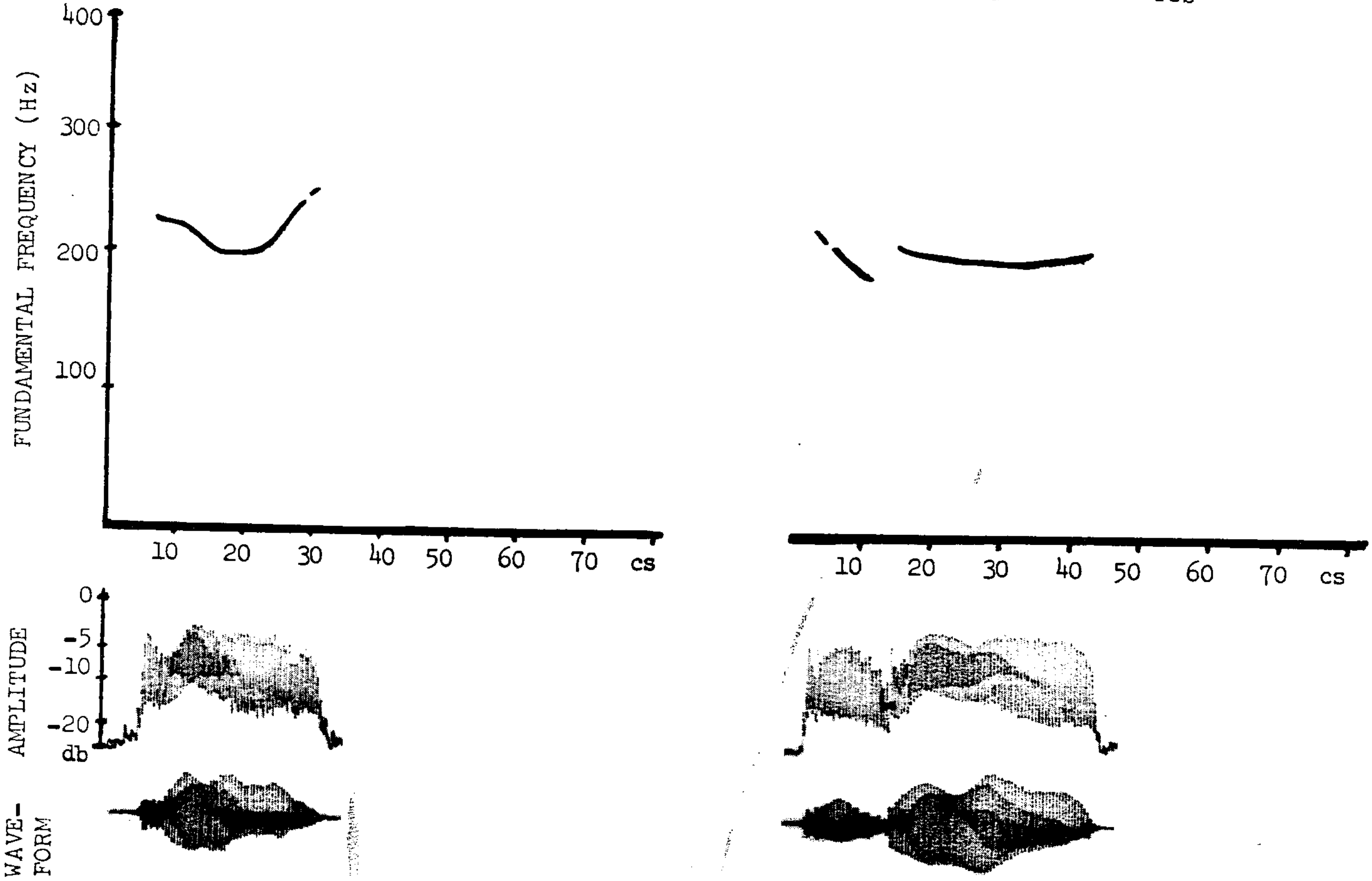


Fig. 5.7 Speaker 2 $\text{Yes}^{\text{bored}}$

C = -62 Speaker 2 $\text{Yes}^{\text{interested}}$

incorrect judgment. This is, of course, extremely effective communication of the intended attitudes. What prevented Speaker 4 from achieving an even higher score was the fact that nine listeners labelled both stimuli as *Yes*$^{\text{bored}}$. This probably resulted partly from her normally laconic vocal style, and partly from the rather long level nuclear portion of her "interested" contour.

Yes$^{\text{agreeable}}$/*Yes*$^{\text{disagreeable}}$

Mean C-score: 65

Mean C-score for top 11 speakers: 78.55

Median C-score: 88

Number of speakers for whom there were

21 or more correct judgments: 10/12

28 or more correct judgments: 10/12

With the exception of the anomalous stimuli of Speaker 10 (Fig. 5.11) and the almost undifferentiated contours of Speaker 4 (Fig. 5.10), there was very effective communication on this contrast. Here again, as on the "bored"/"interested" contrast, the major information seemed to be communicated by the shape of the f_o contour, with the dominant cue being the almost totally flat contour for *Yes*$^{\text{disagreeable}}$, as seen in Figs. 5.8-9. The best contrast for the narrow-ranged "disagreeable" contour was an "agreeable" contour with a very wide range. In the examples shown for Speakers 8 and 6, this contour was rising, but Speaker 1 achieved a very good score (C-94) by contrasting his narrow-ranged "disagreeable" stimulus with a moderately high nuclear-peaked 231# contour for *Yes*$^{\text{agreeable}}$.

The poor performance of Speaker 4 is easily explained by the great similarity of the stimuli she produced, resulting in 39 neutralized judgments; 32 listeners labelled both stimuli "agreeable", while seven heard them both as "disagreeable".

All of the speakers made some slight attempt to supply additional information by means of paralinguistic cues. Once again, these effects were quite subtle (usually consisting of a somewhat hard edge to voice on the "disagreeable" utterance), and probably went un-noticed on a single hearing. Perhaps the most audible of these cues was produced by Speaker 8 (Fig. 5.8), who achieved a "weary" effect on her *Yes*$^{\text{disagreeable}}$ utterance by means of a dropping amplitude combined with a rather "breathy" voice quality.

Just as in the case of Speaker 11's "disbelief" contour (p. 76),

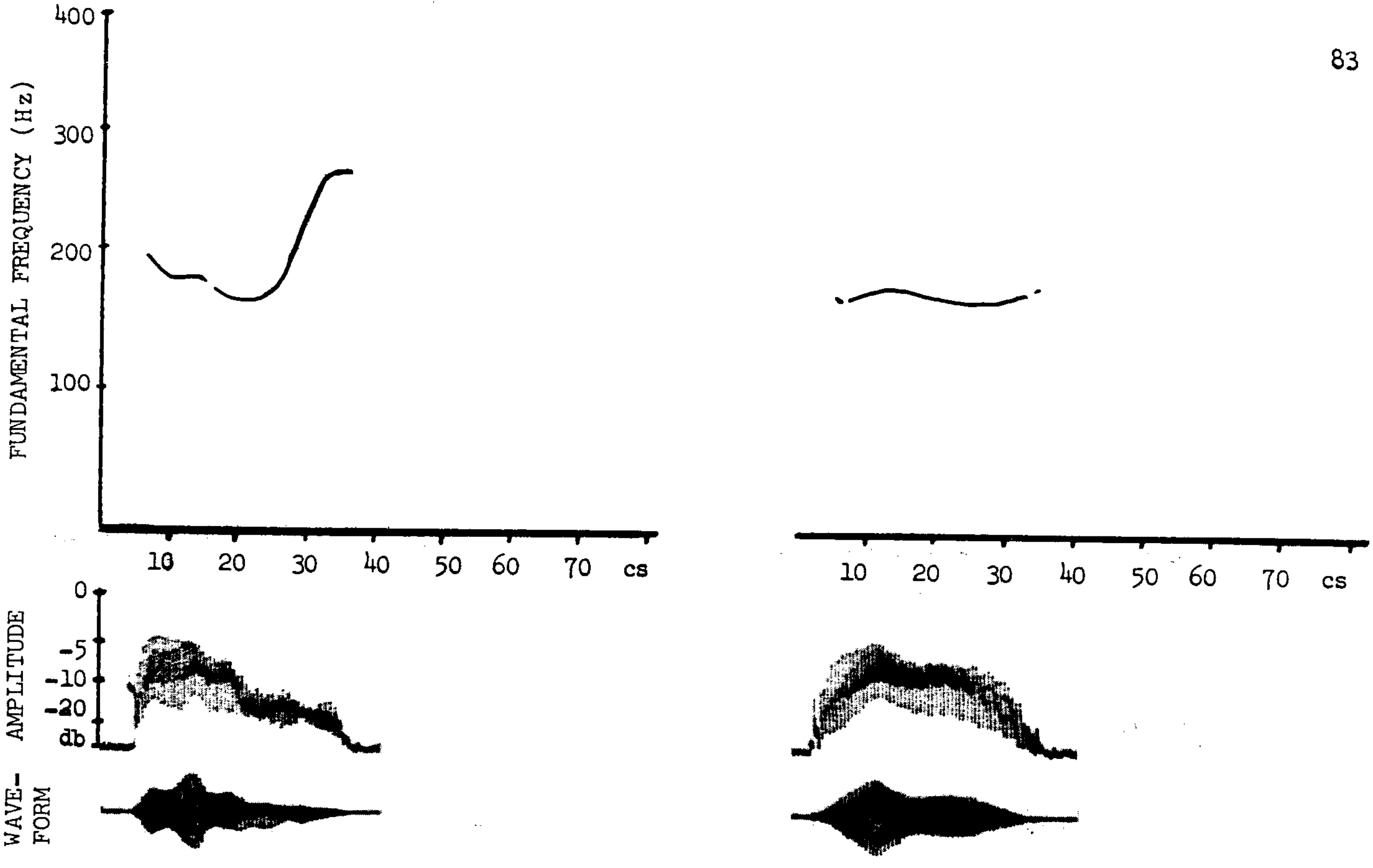


Fig. 5.8 Speaker 8 Yes^{agreeable} C = 100 Speaker 8 Yes^{disagreeable}

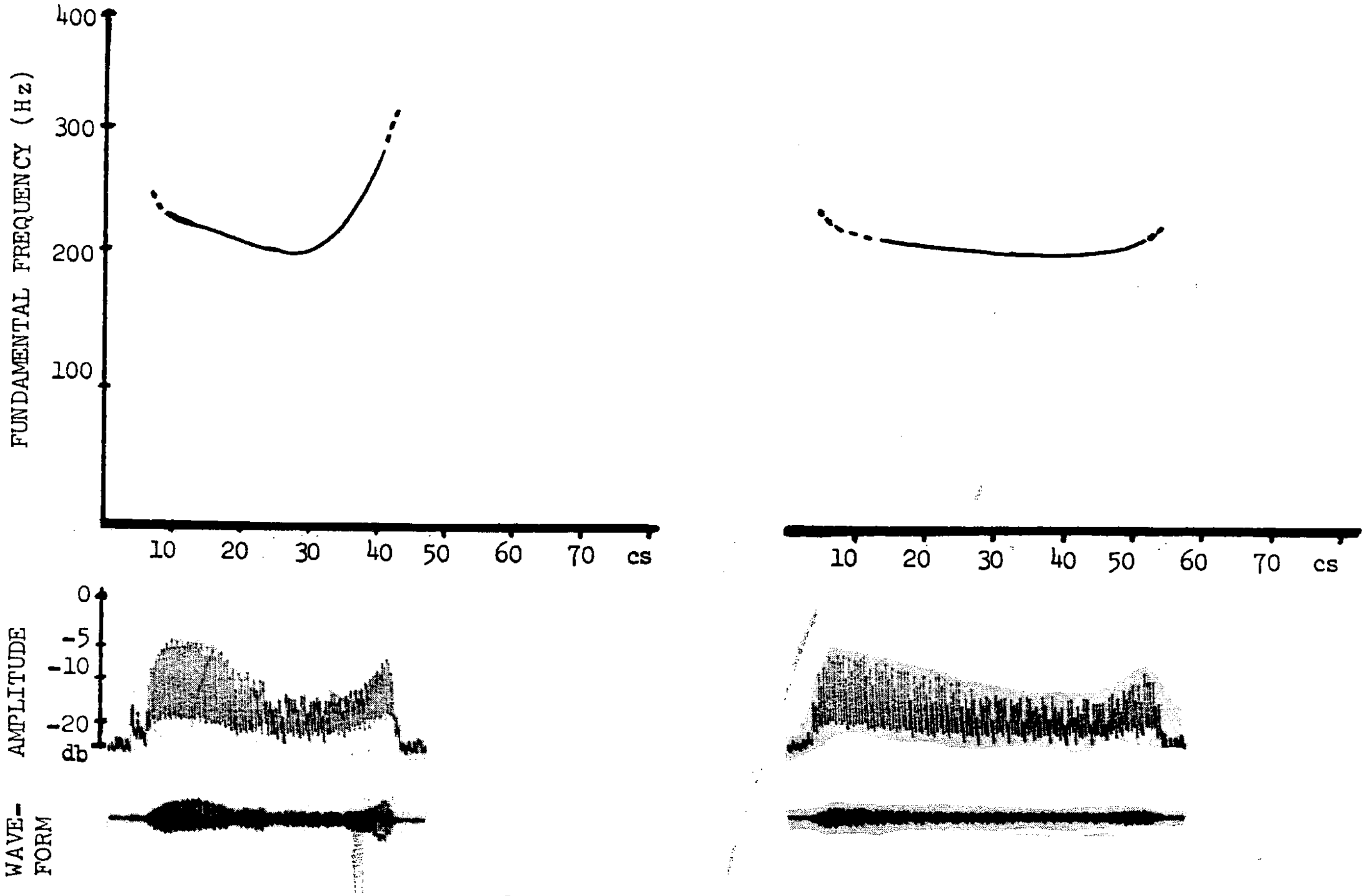


Fig. 5.9 Speaker 6 Yes^{agreeable} C = 98 Speaker 6 Yes^{disagreeable}

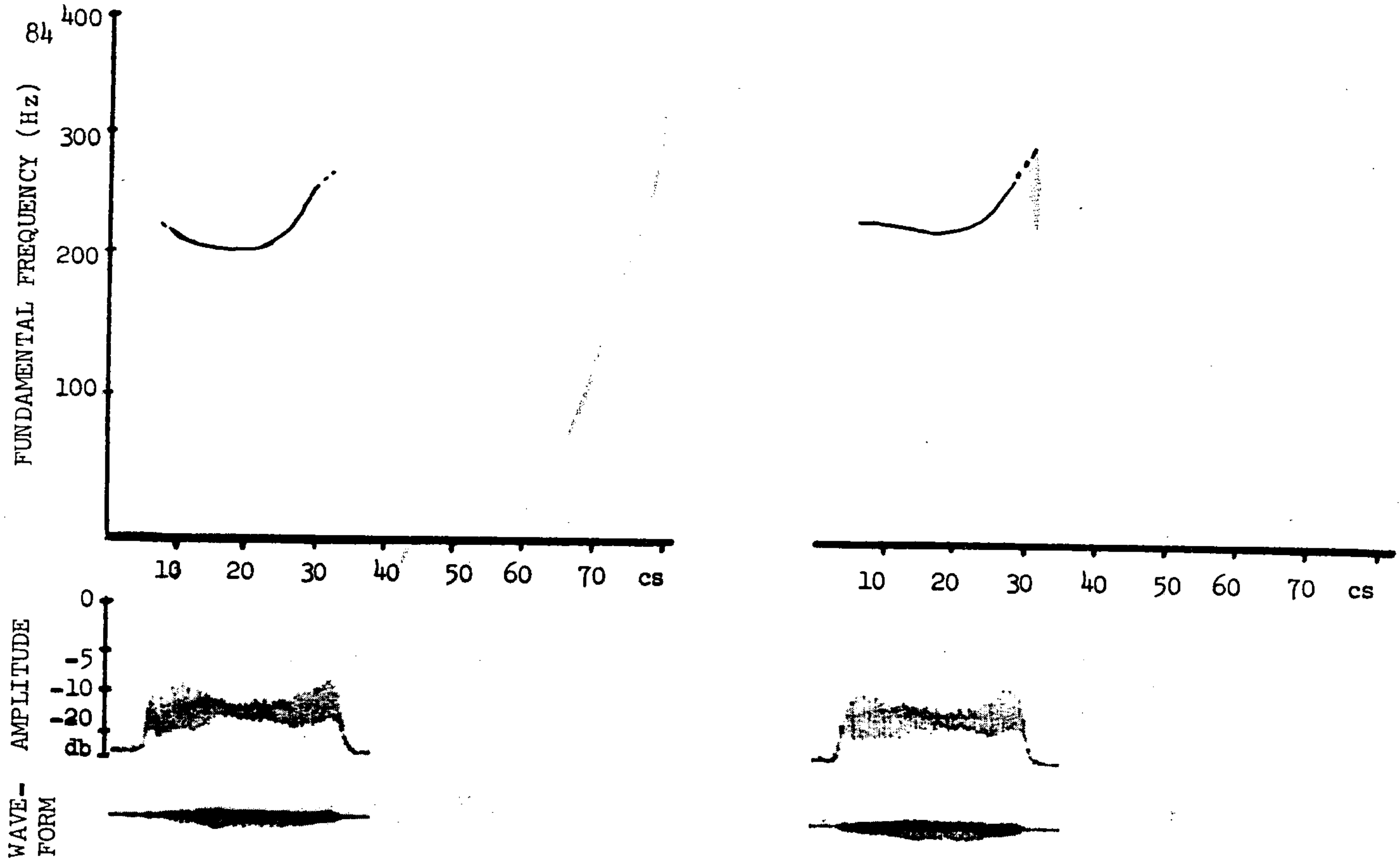


Fig. 5.10 Speaker 4 Yes$^{\text{agreeable}}$ C = 18 Speaker 4 Yes$^{\text{disagreeable}}$

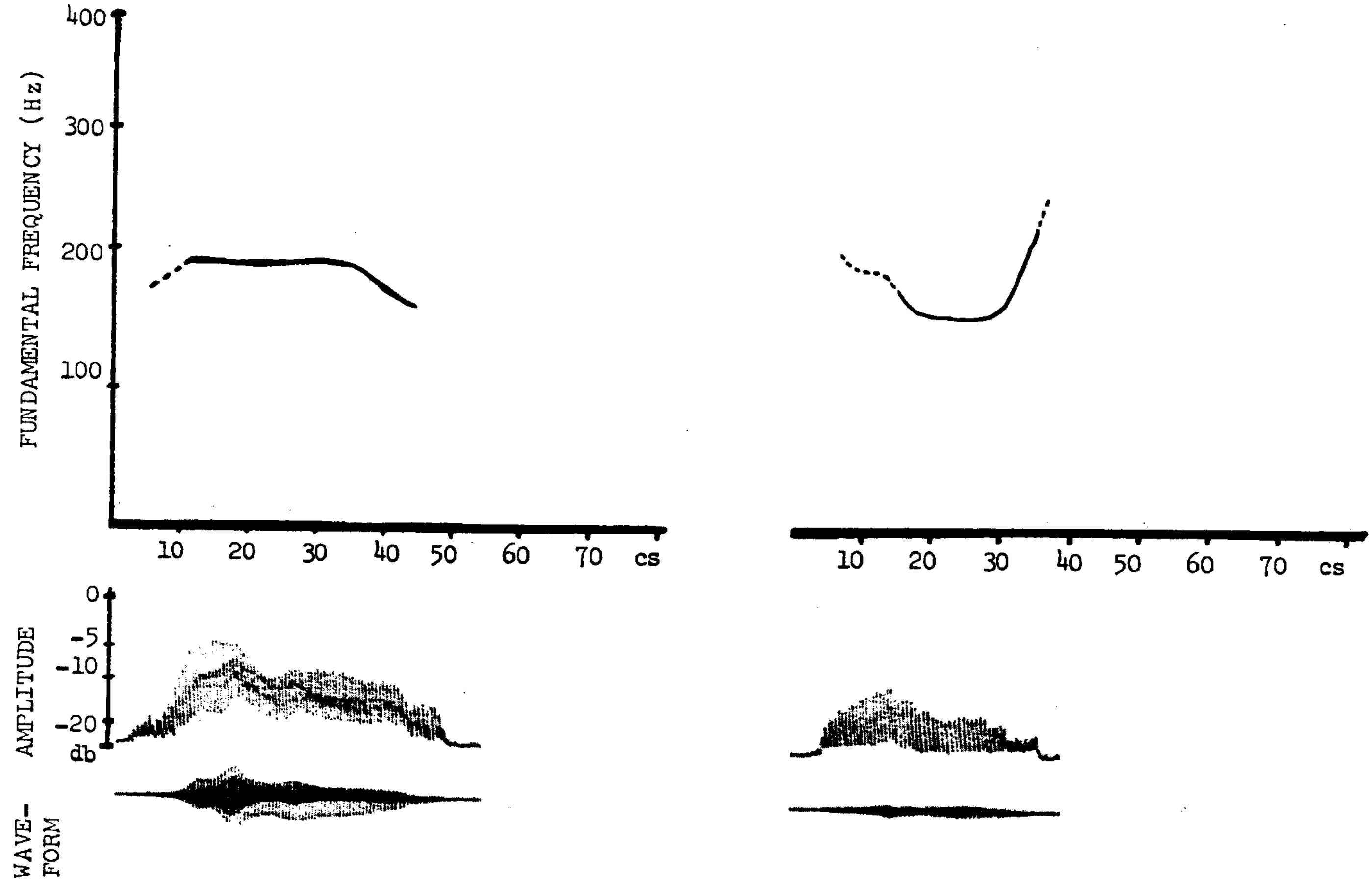


Fig. 5.11 Speaker 10 Yes$^{\text{agreeable}}$ C = -84 Speaker 10 Yes$^{\text{disagreeable}}$

it was the unsuccessful use of paralinguistic cues which led to the anomalous performance of Speaker 10 on this test category. In essence, she tried to make some subtle paralinguistic cues do all the work of communicating these two contrasting attitudes. To her rather neutral f_o contour for *Yes*$^{\text{disagreeable}}$ she added a somewhat tight voice quality and a slight hiss on the final /s/. Apparently both of these cues escaped the notice of the listeners, who categorized this curve as "agreeable". On her intended "agreeable" stimulus, Speaker 10 attempted to communicate a sort of little girl "laughing voice" quality. However, deprived of the visual cues which usually accompany such a paralinguistic effect, the listeners failed to notice the voice quality cue, and apparently responded instead to the flatness of the f_o contour in labelling this utterance as "disagreeable". We seem to have here another case of contour shape overriding voice quality as a perceptual cue for the communication of emotional attitude. However, such a conclusion is no more than speculation, since we did not have the opportunity to contrast the effects of these parameters in a controlled minimal test situation, as in a synthetic speech experiment.

Yes$^{\text{calm}}$/*Yes*$^{\text{angry}}$

Mean C-score: 96.67

Median C-score: 97

Number of speakers for whom there were

21 or more correct judgments: 12/12

28 or more correct judgments: 12/12

The *Yes*$^{\text{calm}}$/*Yes*$^{\text{angry}}$ contrast had the best speaker performance of any test category. Only the two "unemphatic"/"emphatic" contrasts showed performance approaching this level. One consequence of this very high performance level is that there is no way to show a significant difference between the "best" and "worst" performers, since 47 listeners correctly categorized the utterances of Speaker 2, the "worst" performer.

Since even infants are capable of communicating the fact that they are angry, this level of communication was hardly surprising. What was surprising was the variety of means used for communicating anger, particularly as compared with the uniformity of the "calm" stimuli. Taking the latter first, eleven of the speakers expressed the "calm" stimulus with contours of the type shown in the illustrations. Of these, Speaker 2's "calm" contour had by far the widest f_o range. The others more closely resembled the slightly scooped contours with very slight terminal rises shown in Figs. 5.12-13. Even Speaker 12's falling contour might be thought of as a mirror image of these gently rising curves, since hers rose from a precontour of 165Hz to a nucleus of 175Hz, then gently

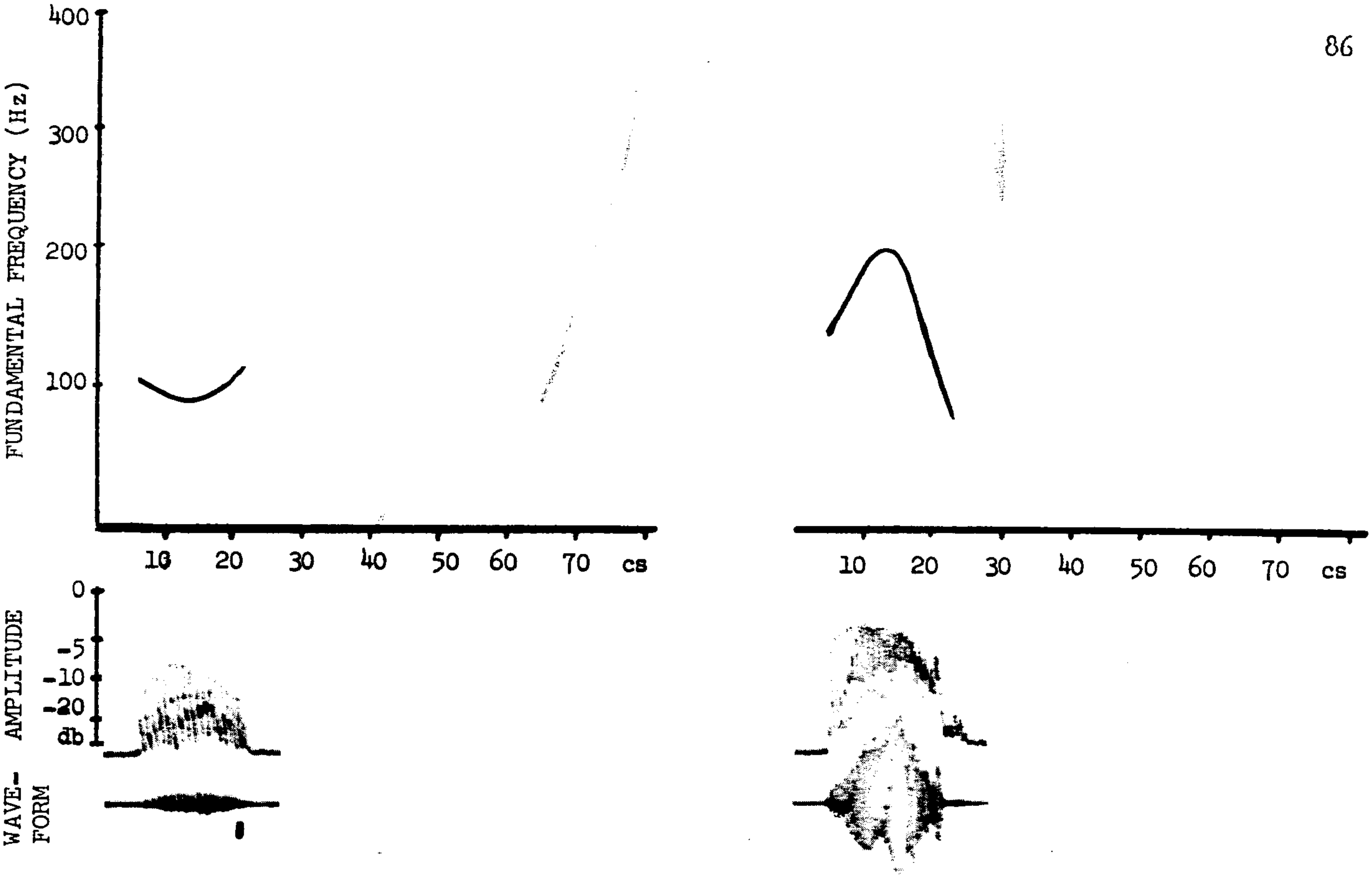


Fig. 5.12 Speaker 9 Yes$^{\text{calm}}$ C = 100 Speaker 9 Yes$^{\text{angry}}$

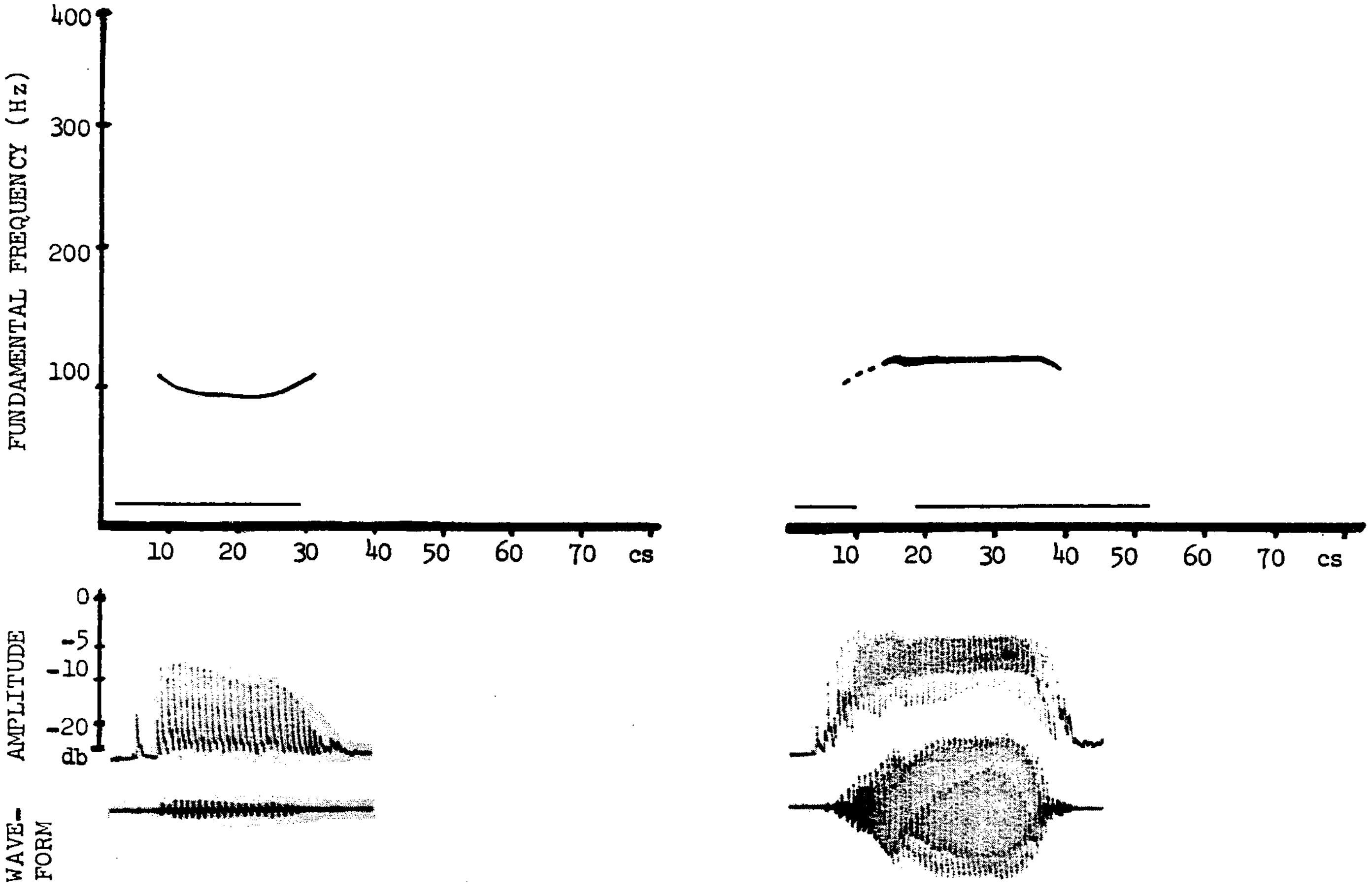


Fig. 5.13 Speaker 11 Yes$^{\text{calm}}$ C = 100 Speaker 11 Yes$^{\text{angry}}$

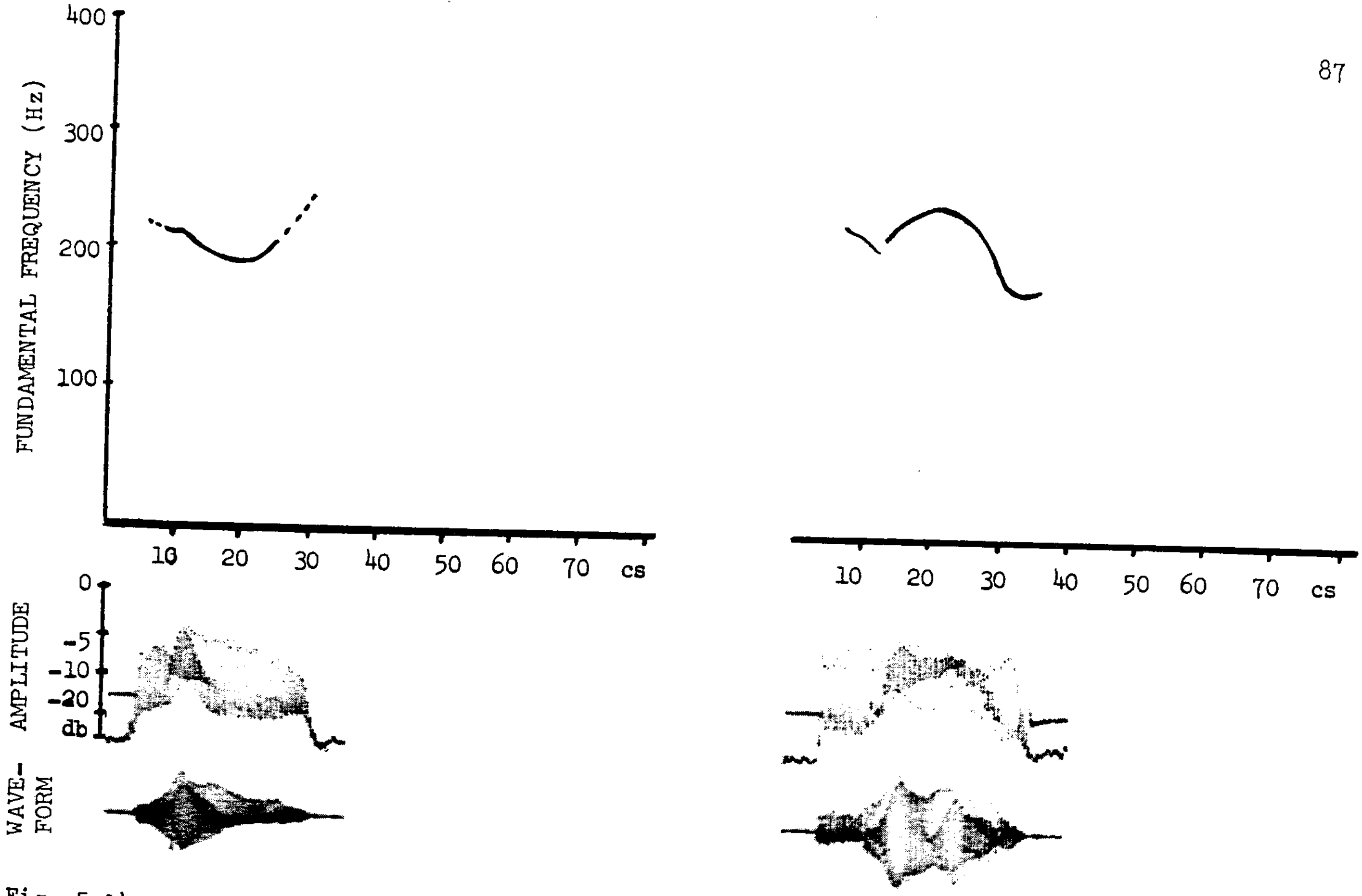


Fig. 5.14 Speaker 2 Yes^{calm} C = 90 Speaker 2 Yes^{angry}

fell to a turning point at 155, followed by a brief level terminal glide. Apparently, the "finality" of a distinct terminal fall, and the "questioning" of a distinct terminal rise were felt to convey attitudes potentially suggesting "anger", so that the speakers chose to express calmness through contours with very gentle, tentative rises and falls.

For communicating anger, almost all of the speakers made simultaneous use of three overlapping cues: a wide-ranging f_o contour (usually with a high nuclear peak and terminal fall, sometimes with a strong scoop to the nucleus followed by a high terminal rise), a large increase in amplitude, particularly at the nucleus, and the paralinguistic cue of hard-edged voice quality. The one major exception to this pattern was Speaker 11, who used a very flat f_o contour, combined with the usual amplitude and voice quality cues. There was one interesting difference between the use of the voice quality cues here and in the other test categories. On the "calm"/"angry" contrast, only one of the speakers made the mistake of spreading the hard voice quality across both stimuli (cf. p. 76 above). The others were able to produce a neutral, or even markedly soft voice quality on the "calm" stimulus, then immediately follow it with a hard (and loud) "angry" utterance. We might regard this ease and naturalness of production as one more sign of the basic nature of this contrast in human speech communication.

Yes$^{\text{angry}}$/*Yes*$^{\text{contained}}$ (anger)

Mean C-score: 78.83

Median C-score: 84

Number of speakers for whom there were

21 or more correct judgments: 11/12

24 or more correct judgments: 10/12

28 or more correct judgments: 10/12

As in the case of the *Were*^{1}/*Were*^{2}/*Were*^{3} contrasts, this test category was deliberately included as a theoretically "difficult" one. Again, it turned out not to be, as can be seen from the figures above, and from the fact that the ten "best" speakers had C-scores of 76 and above.

The shape of the f_o contour, variations in amplitude, and voice quality cues all contributed to the successful contrasting of these two attitudes. Speaker 7 (Fig. 5.15) shows the most easily perceived contrast of f_o contours: the sharply-peaked, high amplitude contour (with an amplitude crescendo at the nuclear peak) which we previously noted for the "angry" stimulus, contrasted with a deliberately flattened f_o contour for the "contained" utterance. At first glance, the amplitude curves for the two stimuli seem rather similar. However, a closer examination shows a distinct difference. The amplitude curve for the "angry" stimulus slopes smoothly downward from the nuclear crescendo, while the "contained" amplitude curve, after a moderate initial rise, is sharply "pinched", and then held constant at its reduced level for a considerable time period. This pinched amplitude curve is seen even more clearly in the "contained" stimulus produced by Speaker 10 (Fig. 5.16), where the depressed section of the curve is followed (after a very slight voice break) by a rise, which highlights the "pinched" section even more. This effect in the amplitude curve is accompanied by a simultaneous "pinching" or "squeezing" of the voice quality, which no doubt aided both speakers in communicating the "contained anger" attitude to the listeners. Most of the speakers produced such a squeezed, tight voice quality on their "contained" stimulus. The use of this squeezed amplitude and voice quality by the "worst" performer, Speaker 6, was rendered less effective by two factors: a slight anticipation of these qualities in the "angry" stimulus, and the great similarity of the f_o contours. These factors, together with the very soft voice production which Speaker 6 manifested on all of her utterances, probably accounted for the fact that 34 listeners categorized both stimuli as *Yes*$^{\text{contained}}$.

Apparently this paralinguistic cue of "squeezed" voice quality is extremely important in communicating the nature of the "contained anger" stimulus. Although we have previously noted cases in which some of our

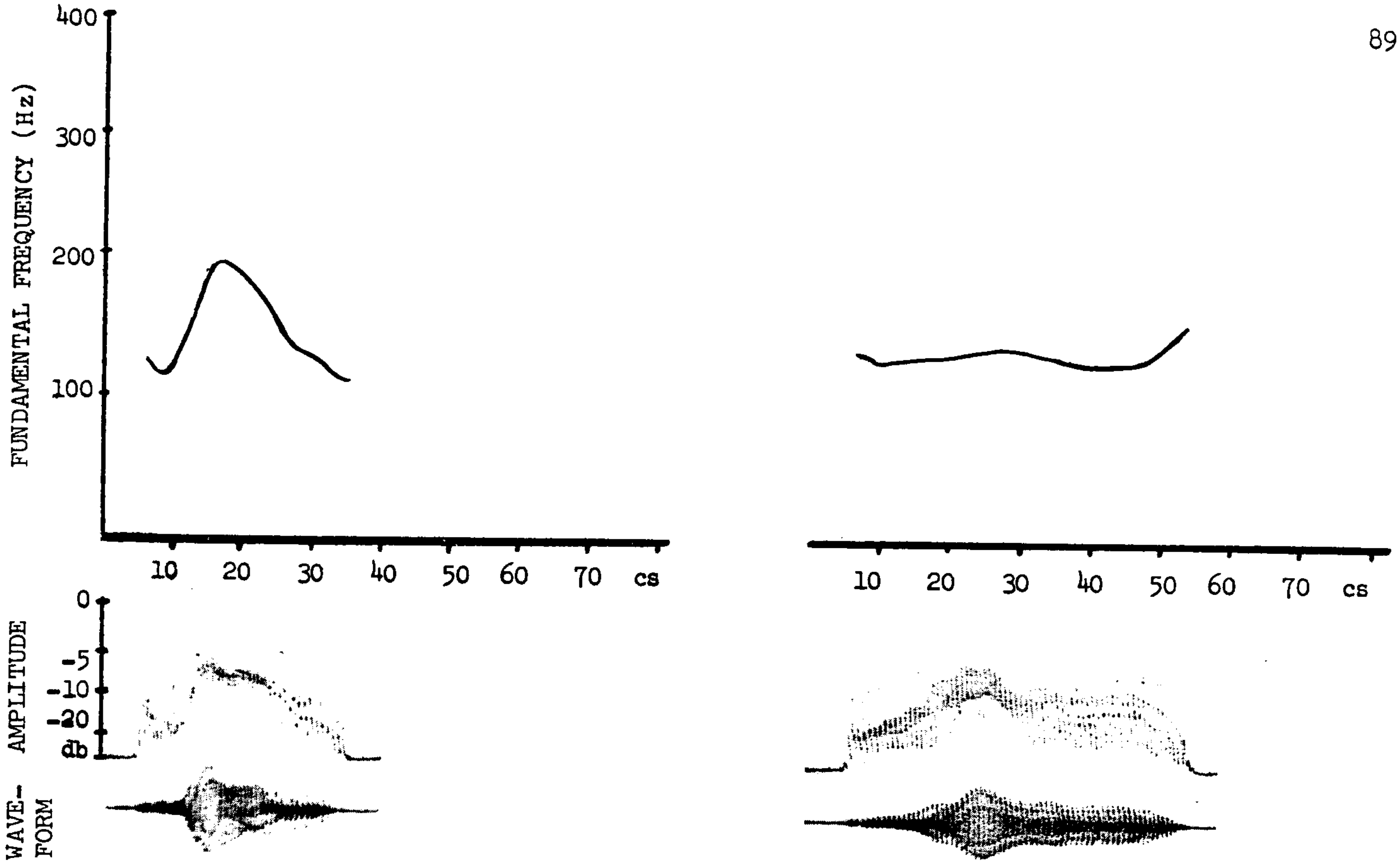


Fig. 5.15 Speaker 7 Yes^{angry} C = 100 Speaker 7 Yes^{contained}

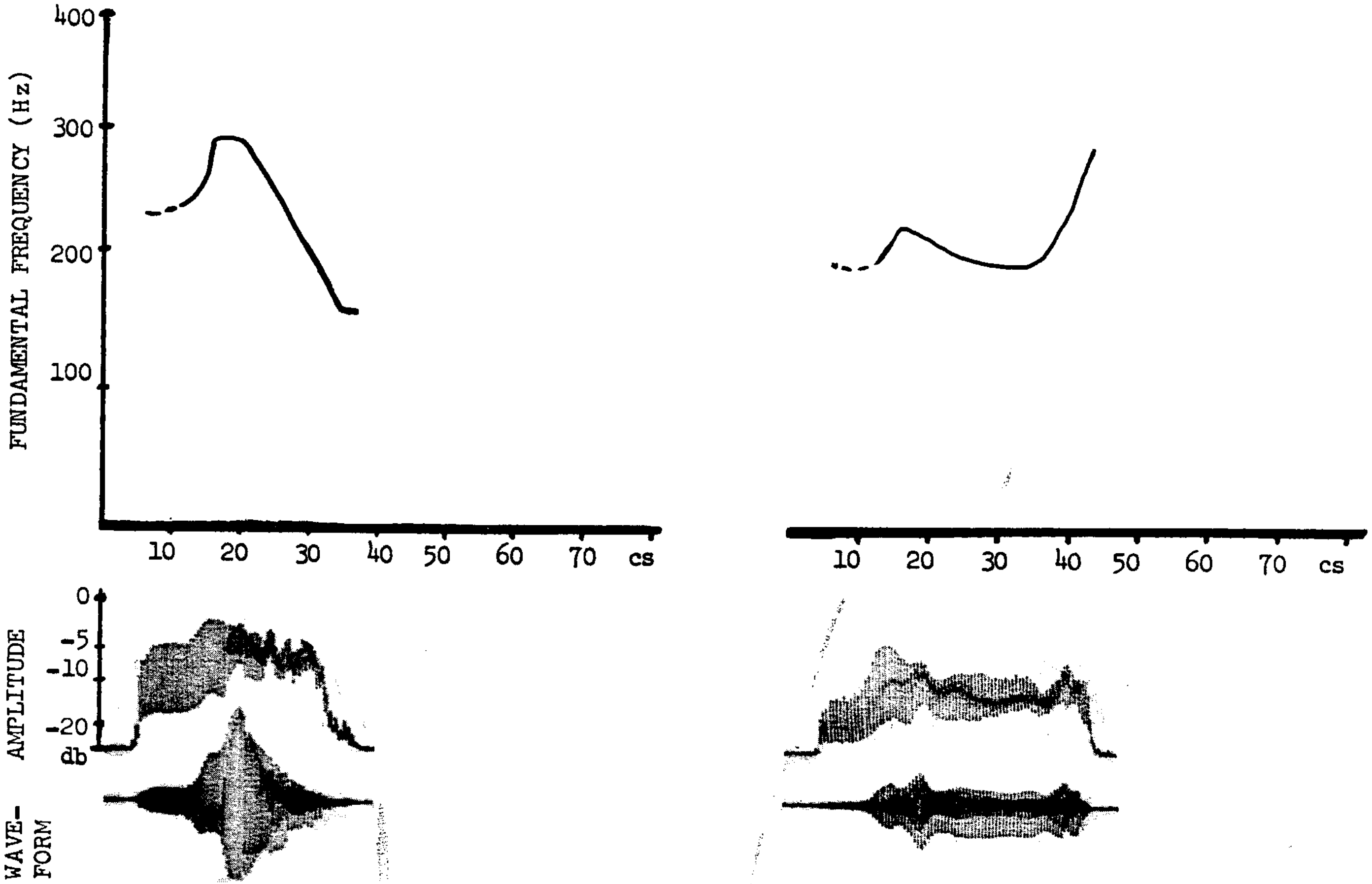


Fig. 5.16 Speaker 10 Yes^{angry} C = 100 Speaker 10 Yes^{contained}

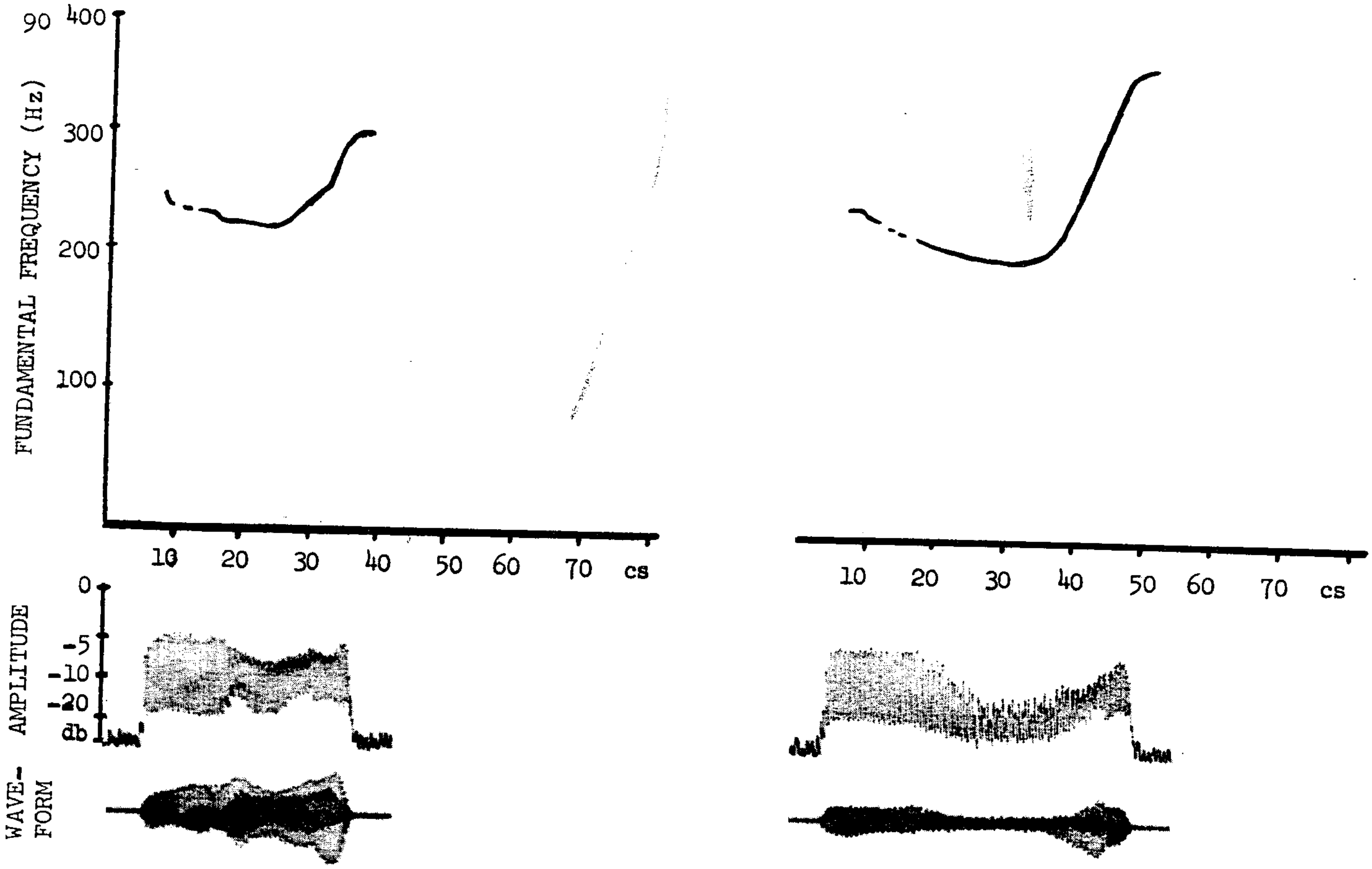


Fig. 5.17 Speaker 6 Yes^{angry} C = 28 Speaker 6 Yes^{contained}

generally very effective "sophisticated" speakers failed to communicate certain contrasts because they depended upon voice quality cues to disambiguate their f_o contours, this category produced cases of the opposite type. Speaker 11 (cf. Fig. 5.13) produced an f_o contour for *Yes*^{angry} which was more typical of the contours for *Yes*^{contained}. He contrasted this unusual contour with a "contained" utterance with a moderately peaked nucleus and a gentle terminal fall. Yet, by successfully contrasting a very tight, "squeezed" voice quality (modulated into creaky voice at the end) on his "contained" stimulus with a clear voice quality (and somewhat higher amplitude) on his "angry" stimulus, he achieved a C-score of 94. Similarly, Speaker 12 produced nuclear-peaked contours for both *Yes*^{angry} and *Yes*^{contained}, but her "contained" stimulus was much longer (52cs versus 34cs), and had a much higher nuclear peak (510Hz versus 265Hz). While the extra length no doubt contributed to moderating the potentially "angry" effect of the very high nuclear peak, her very successful use of a tight, "squeezed" voice quality accompanied by a simultaneous "squeezing" of the amplitude on the *Yes*^{contained} utterance also probably aided in communicating this contrast and achieving a C-score of 98.

Summary of Results for Emotional Test Categories

There were five "emotional" test categories, consisting of contrasts between *Yes*$^{\text{belief}}$/*Yes*$^{\text{disbelief}}$, *Yes*$^{\text{bored}}$/*Yes*$^{\text{interested}}$, *Yes*$^{\text{agreeable}}$/*Yes*$^{\text{disagreeable}}$, *Yes*$^{\text{calm}}$/*Yes*$^{\text{angry}}$, and *Yes*$^{\text{angry}}$/*Yes*$^{\text{contained(anger)}}$. All of these contrasts were elicited by means of cue cards. There was generally very effective communication of these contrasting emotional attitudes, with the "calm"/"angry" category being by far the easiest, and with the "angry"/"contained" category producing better communication than had been anticipated. Paralinguistic cues, particularly with respect to voice quality, played a much larger rôle in these categories than they had in the syntactic test categories. However, with the exception of the "contained anger" stimulus, the shape of the fundamental frequency contour generally seemed to override voice quality as a perceptual cue.

Chapter 6: Some Implications and Extensions of this Study

Functions of Intonation

Throughout our earlier discussion, we centered upon intonation as a means of signalling information of the type largely covered by Hultzen's "syntactic" and "rhetorical" patterns, and by Halliday's (1963, 1967) dimensions of "tonicity" and "tone", i.e., information concerning basic contrasts such as finitive/continuative, unemphatic/emphatic, and what we might loosely typify for the moment as "unemotional"/"emotional". We shall return shortly to a further discussion of these areas of intonation, but first we should look at a more basic area of intonational function -- the use of intonation contours as a means of *organizing* speech communication into units variously described as "tone groups" (Halliday, 1967), "phonological phrases" (Bierwisch, 1966), "phonemic phrases" (Trager and Smith, 1951), and "breath-groups" (Lieberman, 1967). The existence of this organizing function is so basic to speech communication that it is easy to overlook it, and to concentrate exclusively on the higher levels of intonational function, but to do so would be an error, since the higher functions can best be described as growing out of this basic function.

The basic fact here is that an intonation contour seems to serve as a kind of "container" for the surface string comprising a communicative unit, which later corresponds to an underlying sentence. Although the very young infants studied by Bosma, Truby, and Lind (1965) commonly let their cries go past an expiration and into the succeeding inspiration, this practice of breaking the intonational "container" of a message apparently does not survive even into the babbling stage, and the only occurrences in adult speech are to be found in cases such as the report of the Apache massacre gasped-out by the dying messenger ("... came over [gasp] the hill ... killed [gasp] Sam ...") which we all remember from our childhood Saturday afternoons at the movies, and in the occasional use of an ingressive airstream on the first part of the word *yes*, as reported by Abercrombie (1967:25).

Some recent psycholinguistic studies suggest that this use of an intonation contour as a container for speech communication may be rooted in the requirements of the speech-processing mechanisms in the brain. It seems likely that these mechanisms require phrases to be delimited. In his own experimental work, Johnson (1965) found that linguistic phrases tended to hang together, so that subject errors were higher across phrase-groups than within them. These findings suggested to him that language was more than just serial units, and he therefore commended the work of Miller and his associates, which he described as "... an attempt to describe all sequentially ordered behavior in terms of a hierarchy of response units, with higher level units encompassing the lower level units. They have suggested that, on a behavioral level at least, rather than conceive of language sequences as sets of serial dependencies, one can view them as sets of 'nested dependencies' reflecting the units-within-units concept." (Johnson, 1965:474)

Can intonation serve as a means of signalling the way in which items in an utterance hang together? A recent study indicates that it can. O'Connell, Turner, and Onuska (1968) exposed subjects to various strings (including nonsense-strings), presented both in a "monotone" and with "normal English intonation" (the authors give no further specification). They found that "the facilitative effect of grammatical structure was apparent only in the intonated versions," and that "the recall time for intonated presentations was significantly shorter than the recall time for monotone presentations at all levels of structure." (115) They therefore raise the possibility that "... even in the absence of congruent grammatical structure, intonation tends to suggest syllable groupings and thus a strategy for recall." (115)

In his recent book, previously discussed in Chapter 2 above, Lieberman commendably attempted to deal with this organizing aspect of what he calls the "breath-group", and to see whether it might cast additional light upon some examples of potential phonological disambiguation of syntactic strings, such as were studied in the heyday of "phonological syntax." Unfortunately, his attempt is plagued by serious theoretical deficiencies. In his remarks on Stockwell's example *I'll move on Saturday*, with its interesting ambiguity in the verb (*move* versus *move on*), Lieberman insists that the disambiguation would be made by different placements of the end of a marked breath-group (Lieberman, 1967:110-113). But since the breath-groups assigned to the clauses containing *move* and *move on* would both be of the marked [+BG] type, and the only difference would be in their domain (one including only *move*, the other extending to include *on*), we are immediately confronted with a difficulty in that Lieberman has only an *ad hoc* and awkward system for stating the domain of a breath-group (140).

In his recent review of Lieberman's book, Kim (1968) has noted that it is Lieberman's confusing double usage of the term "breath-group" as an abstract formal unit, and as a concrete, physical event (i.e., "expiration") which causes Lieberman to conclude that "the scope of a breath-group can be any constituent of a derived phrase marker" (113), a conclusion which prevents him from giving a systematic set of breath-group assignment rules. To remedy this defect, Kim sketches his own set of generative rules of intonation:

> (1) Assign the feature [-BG] to each occurrence of *S*, and only *S*, in the deep phrase marker (this is to assume that [BG] is an intonational feature whose scope is a sentence, not just any constituent). (2) Change [-BG] to [+BG] by a later rule, if the sentence contains *Q* + *whether* (i.e., if it is a yes-no question), in the way suggested by Lieberman in Chapter 6. (3) Change by another rule every [-BG], except the one that appears with the rightmost *S* in the derived phrase marker, to [+BG]. (4) Convert, by rules of synthesis, [-BG] to a falling pitch and [+BG] to a not-falling pitch. (Kim, 1968:839)

In a footnote, Kim adds the suggestion (apparently based on Stockwell, 1960) that

> ... the number of [BG]'s in the derived phrase structure probably corresponds to the number of *S*'s that survived to the surface structure. That is, it seems likely that if a node *S* in deep structure becomes deleted in the course of a derivation, e.g. by a "tree-pruning" rule as suggested by Ross 1966, and is not present at all in the derived phrase marker, then the feature [BG] will also have been deleted with it. Thus, in sentences like *Don married a pretty girl* from *Don married a girl* (S_1) *who is pretty* (S_2), or *I saw Ralph and John* from *I saw Ralph* (S_1) *and I saw John* (S_2), it is likely that there is only one [BG] associated with each sentence, the other [BG] having been deleted with the deletion of the node *S* in the course of the transformational derivation. (839)

Kim's rules go some distance toward making the notion "breath-group" *syntactically* meaningful, and would, we believe, be adequate for dealing with another of the "phonological syntax" examples. *They decorated the girl with the flowers,* since the differing deep structure trees for this string could be made to signal the appropriate breath-group division. But they would not handle our original example of *I'll move on Saturday*, since information from the stress cycle is crucially important here. It would therefore appear that rules making possible contrasting phonological realizations of the different deep structures underlying strings such as *I'll move on Saturday* require an exacting marriage between intonation rules of the general type proposed by Bierwisch and by Kim, and inherent stress cycle rules of the type proposed by Chomsky and Halle (1968). Such a marriage will not, however, be consummated within these pages.

The importance of stress phenomena in their interaction with intonation contours can be seen in the serious failure of Lieberman's *Sprachgefühl* in respect to another of his examples:

> Consider the string of words *I saw the boy who fell down the stairs*. If these words were uttered on one normal breath-group, the listener would hear the sentence *I saw the boy who fell down the stairs*. The underlying phrase marker of this sentence, which contains a relative clause, is similar to the underlying phrase markers of the two simpler sentences *I saw the boy* and *The boy fell down the stairs*. The intonation indicates that the string of words *I saw the boy who fell down the stairs* constitutes a complete sentence. If the same string of words is uttered on two normal breath-groups
>
> [*I saw the boy*] [*who fell down the stairs*]
>
> the listener will treat the speech signal as though two sentences were uttered, the simple declarative sentence *I saw the boy* and the interrogative sentence *Who fell down the stairs?* The underlying phrase marker of each sentence will be derived independently ... The presence of the first marked breath-group is thus very important if the speaker wants to break up the sentence *I saw the boy who fell down the stairs* into two breath-groups, since the listener has no other way of telling (fn.: "In the absence of other context, such as previous parts of the conversation") that the two strings of words [*I saw the boy*] and [*who fell down the stairs*] are not independent sentences. (168-69)

But all of this pontification is surely on the wrong track, since Lieberman has ignored the vital matter of potential for pitch accent (cf. p. 7 above). Vanderslice (1969:4) would appear to be correct in rejecting Lieberman's above example on precisely these grounds:

> No doubt the use of two falling intonations would be unusual in pronouncing these as a single sentence, but there remains a criterial difference of accentuation: interrogative *who* is canonically accentable; relative *who*, canonically weak (cf. Vanderslice, 1968, pp. 53-72).

Despite these failures, Lieberman deserves credit for attempting to re-focus attention upon this significant area of intonation function.

In discussing the use of finitive versus continuative intonations, we are, of course, precisely in the area of intonation called "syntactic" by Hultzén, who gave us both a better example of this problem and a more reasonable assessment of the communication situation:

> An even simpler case occurs when the speaker ties together a loose sentence such as *He'll get in trouble, if he does that.* The first clause is completion in text shape. The speaker knows, however, that the qualification is coming and holds the ending

> up by reversing the text-appropriate closed tune. What makes this reversal significant is that it is not obligatory. Some speakers don't do it, and we say of them that they are difficult to follow. (Hultzén, 1961:660)

In saying that those speakers who do not make adequate use of the intonation system's signalling capacity are "difficult to follow", Hultzén seems to have come close to the essential function of intonation, which we believe is that of making the speakers of a language "easy to follow". This overall function is divided into a least four sub-functions: (1) the use of an intonation contour as a "container" for the message. This appears to go beyond what is envisioned in Hultzén's description of his "formal pattern" (cf. Chapter 2) as merely the organization of accents and unaccents, but since this use of intonation appears to be automatic at a very early stage of the child's acquisition of his native language, we shall add no more here to our comments earlier in this chapter on this function. (2) The use of an intonation contour as a means of signalling syntactic information. Although we earlier criticized Hultzén for limiting his "syntactic pattern" of intonation in English to the signalling of continuity versus discontinuity (our Chapter 4 shows that finer grades of communication are possible through intonation alone), it is true that the complete/incomplete contrast is a basic one, which underlies other contrasts. Stockwell (1960) attempted to emphasize this fact by writing a choice between "Discontinuous" and "Continuous" contours into his base rules. Somewhat later, Bolinger expressed the same basic notion:

> Unfinished business, besides telling us that we are in the middle of an utterance, next transfers the high pitch of the middle to the end, enabling us to leave things like questions deliberately unfinished for the interlocutor to finish them. A language that uses high terminal pitch for unfinished business is like English whether or not it does so for questions; questions are secondary. (Bolinger, 1964 a:843)

In looking for universal aspects of intonation, it is necessary to ignore many surface details and concentrate upon basic elements such as the finitive/continuative distinction. In doing the phonetic study discussed in the preceding three chapters, we found it useful to hug the phonetic ground fairly closely. Any attempt at a full-scale theory will, presumably, have to incorporate both kinds of information.

We shall return to syntactic uses of intonation when we discuss the development of systems of intonation in the child and adult. Now we shall turn to the enumeration of the next of our sub-functions of intonation: (3) the use of an intonation contour to signal rhetorical emphasis. This function (Hultzén's "rhetorical pattern") is rather simple from a generative standpoint, resulting from a late EMPHASIS transformation. From a perceptual standpoint, it seems also quite simple, since our listeners had no trouble at all in correctly categorizing "unemphatic" and "emphatic" utterances (see Chapter 4 above). However, from a phonetic standpoint, it seems necessary to observe that the placing of emphasis upon a polysyllabic word or phrase can result in elaborate distortions of the normal enunciation (see p. 69 above).

We turn now to the last of our sub-functions: (4) the use of an intonation contour to signal emotional attitudes. In his paper dealing with universal traits of intonation, Bolinger stressed a conception of intonational schemes growing out of basic emotional metaphors, the chief of these being a tension-relaxation dichotomy. He stresses the emotional underpinnings of intonation elements in passages such as the following:

> An accent language employing relative heights may distinguish old from new or topic from comment, with intonation getting a foothold in the syntax. But the foothold is with one foot; the other one is back there doing its primitive dance ... It is impossible to separate the linguistically arbitrary from the psychologically expressive. Even so simple a thing as a terminal fall shows by its gradience that what counts is how positively through we are... (Bolinger, 1964 a:844)

A somewhat contrary view of emotional aspects of intonation is expressed by Abercrombie (1967:9), who stresses the linguistically conventional nature of such usages. After discussing the effects of fatigue, over-consumption of alcohol, and nervousness on speech, he says:

> More interesting are those indices that do not have a direct physical cause, those from which we infer feelings such as amusement, anger, contempt, sympathy, suspicion, and everything else that may be included under "tone of voice." Indices of this kind are probably more commonly learnt from other people than is generally believed: they are for the most part conventional rather than instinctive expressions of mood and emotion.

There would appear to be truth in both these viewpoints. Bolinger's wide-ranging system of intonational metaphors laid upon other metaphors yields a framework making possible one kind of universal study of intonation. Yet one's attempt to effectively communicate a particular emotional attitude through intonation will certainly be facilitated by having studied the typical intonational means employed for communicating that attitude within that speech community, so that Abercrombie's emphasis upon language conventions is warranted. However, when we try to reason why certain conventions are followed, i.e. when we turn from training speakers to analyzing a community's language habits, we may be right back with Bolinger's attempt at a cultural-anthropological analysis of intonation usage:

> Take a foreign speaker who misuses an intonation and is misunderstood. The naive cultural relativist looks at this and takes it for proof of an accidental similarity in form only. But this is giving up too easily. It fails to make a distinction between meanings and values. For example, a low-pitched fall in two languages may mean finality in both, but finality may be frowned upon sometimes in one community but approved in the other... (1964a:842)

In our experiment, we separated those items which were primarily syntactic from those which were primarily emotional in communication. It has frequently been the case, however, that students of intonation have allowed the observation of affective elements to cloud the study of syntactic communication in intonation. We believe that this has been especially the case in investigations of intonation in child language.

Systems of Intonation in Child and Adult Language

It has generally been assumed that intonation is the first part of the native language to be learned. In this section, we shall argue that it is first learned *and last learned*, i.e. that the learning of intonation begins before the learning of such other language elements as segmental phonemes and syntactic rules, continues through the learning of these other lements, and *may* continue beyond the learning of all parts of the language, except for the addition of lexical items. However, we should note that the learning of *all* aspects of the language probably takes longer than has been supposed (cf. Hunt, 1965). Furthermore, current transformational theories of language change (Halle, 1962, Postal, 1968) require the possibility of the addition of late rules even for segmental elements. Therefore, it is quite probable that the kind of extended acquisition we are arguing for intonation is really the case for all parts of the grammar. However, one difference would still remain, in that our conception of the late intonation rules as consisting essentially of directions for splitting previous intonation categories makes these late rules quite powerful.

In making our claim that the acquisition of intonation takes place over an extended period of time, we are forced to reject much that has been previously written on the subject of child language acquisition. Easiest to reject are the unsupported claims that the child has learned the intonation system of his language by a certain tender age, as in Smith's review of Jassem's *Intonation of Conversational English:*

> Chapter 5 *The tonal unit*, and Chapter 6, *The tune*, endeavor to systematize Jassem's observations. Here the welter of terms and classifications -- 'nuclear tunes', 'prenuclear tunes', 'high falling-rising', and so on -- staggers the reader and makes it impossible for the author to systematize what is really a

> simple and orderly set of linguistic phenomena. As one goes over this portion of the book, it is hard to realize that *a normal child controls the structure of these phenomena before he is two and a half years old.* (Smith, 1955:153; emphasis added)

A more subtle version of this same viewpoint is expressed by Joos in the midst of an otherwise excellent article on language and usage:

> The native language is learned in stages, each stage completed while the next is in progress. The first stage is the learning of the complete pronunciation-system, and normally the books are closed on that before schooling begins. The second stage is learning the grammatical system; this begins about one year later than the first stage began, and it is complete -- and the books are closed on it! -- at about eight years of age. It is not normal to learn any more grammar beyond that age. (Joos, 1964:205)

Joos goes on to say that native speakers who did not acquire constructions such as "to err is human", or the use of the past-perfect tense, missed the boat at an earlier age: "It appears that any who had not learned it by age eight were destined never to learn it, for after that it was too late."

One could object to this statement on various grounds, including the lack of empirical evidence or the fact that Joos has lost sight of the distinction between colloquial and literary English (a surprising slip on the part of the author of *The Five Clocks*!); i.e. "to err is human" and the use of the past-perfect tense both belong to literary English, and are, we believe, usually acquired during the study of literature at the junior high or high school level. But most objectionable is the notion that "the books are closed" on a particular phase of native language learning at a fixed age *for all speakers*. We shall, through the use of our own experimental data and that of Gleitman (1967), argue that *some* speakers keep their books open longer than others, and that this process of differentiated acquisition exists particularly in intonation.

However, we must first take a critical look at the kind of evidence usually brought forth in discussions of infant and child intonation. Let us begin with this general comment by Fry:

> One of the aspects of speech that the child learns to reproduce successfully quite early is the intonation of what is said to him. This is not because rises and falls in pitch are particularly easy to imitate but rather because intonation is closely linked with the affective side of speech; its use grows naturally out of the expressive sounds the child has been making, and the emotional tie between mother and baby ensures that the baby will readily imitate the mood and tone of the mother. (Fry, 1966:191)

It is perfectly true that even an infant can adapt his fundamental frequency range to that of an adult speaking to him (*not* just the mother!) Jakobson noted this fact long ago (1941; English version 1968:24), and Lieberman (1967:45-46) reports an experiment which showed that a 10-month-old boy and a 13-month-old girl adapted their f_0 ranges while babbling to more nearly approximate the voice range of the father or the mother holding them and talking to them at the time. But this imitative use, even as it extends to the imitation of a rising or falling contour, is not what is important in determining the acquisition of a language element, as can be seen from the study by Fraser, Bellugi, and Brown (1963) of the control of grammar by three-year-olds. They treated imitation and production as different skills, and concluded that imitation is "perceptual-motor performance", which does not work through the meaning system. Therefore, what matters in intonation, as in other language elements, is evidence of *contrastive* usage.

Some studies of infant language (cf. Velten, 1943; also Templin, 1957) make no mention at all of suprasegmental features. When one examines those which do include comments on intonation, one finds that the matter of contrastive usage is usually either sidestepped (cf. Lenneberg, 1967:279) or ignored (cf. Grégoire, 1937:74-77; Lewis, 1951:114-16; Jespersen, 1922:111-12; McCarthy, 1954:521-23; Carroll, 1961:337). The typical comment in these works refers to gross reactions of infants to very broad affective (or even paralinguistic) elements in the intonations of adults. Thus Lewis observes that

> ... perhaps as early as his fourth month, he may utter sounds expressive of contentment when someone speaks pleasantly to him. About this same time, or perhaps a little later, he will begin to show distress and cry when he is spoken to sharply — whether 'in reality' or 'in play'. (Lewis, 1963:27)

Certainly these reactions to exaggerated voice quality cues have only the most tenuous connections with the development of a real intonation system. These connections are made even more distant by the fact that one of the closest observers of infant language noticed no examples even of verified imitation:

> A word should be said about intonation in the pre-speaking stages. In the very first diary entry it is stated for *0:1* that Hildegard disliked loud speaking, did not pay any attention to the sounds of words, but was receptive to the emotional appeal of timbre and intonation. I observed nothing worth noting during the babbling stage. More specifically, *there was no instance of imitation of adult sentence intonation* carried by meaningless babbling sounds, which is so often reported in the literature *and which I have myself observed with other children.* Some writers assume that this phenomenon is general. This belief is disproved. (fn.: "It was not observed with Karla either.") (Leopold, 1939: 256; emphasis added)

It is an interesting fact that Leopold did not find imitation of adult sentence intonation during the intensive study of either of his children, but did note such imitation in other people's children, whom he observed on a much more casual basis. This fact suggests that many examples of so-called imitation of adult intonations by infants may be quite accidental in nature.

Imitation of adult sentence intonation does occur among older infants who have reached the early stages of real language learning (as opposed to babbling). But even here there is no evidence in favor of the existence of contrastive usage. The belief that there is such contrastive usage seems to spring from the fact that infants frequently use rising contours. Adults, knowing that such contours contrast with falling contours in adult intonation systems, apparently assume that the rising and falling contours are in contrast in the infant's system also. However, as the Czech linguist Ohnesorg has observed (Weir, 1966:157), most utterances addressed to a child are questions, so that the child's usage of a rising intonation contour remains structurally ambiguous for some time. Weir cites Ohnesorg's observation as a means of resolving a contradiction in her own work, namely "the apparent early use of intonation patterns on the one hand and my own inability to find systematically contrasting patterns with a two-and-a-half-year-old child on the other". (Weir, 1966:158) This influence of adult intonations upon infant imitation choices was noticed earlier by E. G. Pike, who decided to experiment upon her second daughter, Barbara. When Barbara was ready to graduate from babbling to the learning of English (Pike does not give her age at the time), she exposed her only to falling intonations on all lexical items constituting one-word sentences (e.g. *Baby*, meaning "This is a baby."). Barbara used only a falling contour. Later, when left for four days with a family which used the typical baby-talk rising contour, Barbara started using rising contours. Back home again, with words such as *Baby*, which she had heard both in her own home and the other family's home, she would use either a rising or falling contour, whichever was modelled for her by her parents. But with the word *Daddy*, which she had heard only in her own home, she kept using only a falling contour. (Pike, 1949: 21-24) We thus see that even a high level of imitation does not imply the existence of contrastive usage of intonation contours.

Miller and Ervin's study of child grammar acquisition provides some support for this view:

> Children are good mimics of prosodic features, particularly pitch, and they can give the impression of having the pitch-stress system under control. This may be true from a phonetic, perhaps even phonemic standpoint, but does not necessarily entail the use of the prosodic features in the grammatical system. (Miller and Ervin, 1964:28-29)

In her study of the pre-sleep monologues of her two-and-a-half-year-old son, Weir (1962) provides further empirical support for some of the views presented here:

> Whether we follow a description of intonation in terms of pitch levels and junctures ... or whether we adopt an analysis by contours ... it seems impossible for our data to identify the functionally-significant unequivocally. Roughly, there are three pitch levels, *but they are not used contrastively.* A fourth level occurs, higher than any others, in calls and urgent requests. This pitch we can classify best as part of the emotive function of language. In terms of contours, the most frequent one is falling; next in frequency is a rising contour; a sustained one is found least frequently. But here again we were unable to discover a functional relationship among them as we do in standard English. *As a matter of fact, to take an utterance and assign it a certain meaning on the basis of standard English intonation is most misleading.* (Weir, 1962:29; emphasis added)

This emphasis on the non-contrastive nature of those elements which we usually think of as comprising the intonation system of English becomes even more significant when we add it to Weir's following observation:

> Nevertheless, intonation does perform a certain function, but on a different level: *it serves as a marker of sentence boundaries.* The syntactic structures found in the corpus are varied, but a sentence can most readily be defined by an intonation contour with either a final fall or final rise, or by a sustained pitch, each followed by pauses of varying length. The relative length of the pauses becomes particularly significant in our discourse analysis, and the sentence-final pauses are quite consistent in length. There are few instances where a very short pause occurs within the sentence as we have just defined it. The intonation contour is then interrupted by it, and completed after the pause. Some instances of such an occurrence would be:
>
> $\overline{\text{all through}}$ | $\overline{\text{all done}}\searrow$
>
> $\overline{\text{then first lunch}}$ | $\overline{\text{then office}}\searrow$
>
> $\overline{\text{not yellow}}$ | $\overline{\text{red}}\searrow$
>
> (Weir, 1962:29; emphasis added)

When we add to these facts Weir's observation (28-29) that her son Anthony had, from the end of his first year, been able to handle the purely imitative aspects of intonation quite easily, imitating both rising and falling contours, and yet note that some 18 months later he still did not use these contours contrastively (but see our further comment on p. 104), we are then able to add Weir's description of Anthony's system to the few other worthwhile descriptions of infant intonation to suggest the following stages of development:

I. During the first few months of life, the infant reacts in a broad way (through generalized signs of pleasure or displeasure) to polar attributes of voice quality (very soft and pleasant versus very harsh or angry, including pretended anger). This dominance of voice quality as a perceptual cue raises the possibility of a primitive type of analysis-by-synthesis. It may be that the infant is aware that he produces such harsh, strained voice quality through a vigorous tensing of his vocal tract muscles, and does this only when he is angry, and therefore concludes that the adult speaker must be angry. However, since this hypothesis would be very difficult to test in a controlled manner, it must be accorded the suspicion due all unverified hypotheses.

II. During the latter months of the first year, the infant learns to respond to a few very general utterances from an adult, frequently of the type "Baby clap hands!" (often accompanied by gestures displaying the response desired from the infant). Lewis has used an example of this type from Schäfer to argue that "at an early stage, the child shows discrimination, in a broad way, between different patterns of intonation." (Lewis, 1951:115). However, the example deserves closer scrutiny: An infant nine months and nineteen days old responded to the request *Mache bitte bitte* only when it was uttered with the exaggerated type of intonation frequently used with infants, but by the age of 10 months, he responded correctly when the phrase was uttered in an "ordinary" tone of voice. Furthermore, the child gave the same response to *kippe kippe* as to *bitte bitte*, when uttered with the original exaggerated intonation, but not to *lala lala*.

The ability of the child to respond to an "ordinary" tone of voice tells us little or nothing about the nature of his internalized intonation system. Once a stimulus-response pattern has been firmly established, one or another parameter of the original stimulus can be lessened in intensity without decreasing the effectiveness of the response. But this does not prove that the parameter which has been weakened is the dominant one, which is the conclusion reached by Lewis. In assessing this aspect of the perceptual situation, and the matter of the *(Mache) bitte bitte/kippe kippe/lala lala* interrelationships as well, we may benefit from the perspective of Galunov and Chistovich (1965:362) regarding infant comprehension of speech:

> If the set of possible spoken messages and, accordingly, the corresponding responses to them are very limited, each message may be regarded as an individual conditioning signal, and each response as an individual conditioned reaction. Then the problem of understanding the message can be reduced to the problem of recognizing the pattern of this message, which falls well within the scope of the statistical theory of pattern recognition...

> In order to acquire the capability of recognizing a limited set of spoken messages, the discriminating instrument (man, animal, machine) is in no way required to decompose the message into its elements. It is necessary and sufficient to work out the pattern of the message, i.e., to pick out the set of useful criteria by which dissimilar messages are distinguished from one another and to establish the criteria by which decisions are to be made. If the set of possible messages is very small, the most diverse properties of the signals may prove to be useful.

The failure of *lala lala* to elicit the desired response shows that segmental distinctive features (Jakobson, 1941; Jakobson and Halle, 1956) are functioning as cues in this stimulus-response sequence. It is, of course, quite probable that suprasegmental features are also cues here, but to assume (without reliable experimental data) that they are the dominant cues, or even that they closely resemble the intonational features operating in the adult language systems, would be nothing short of presumptuous.

III. Near the beginning of the second year of life, as babbling gives way to the beginnings of real language, the infant learns to utter one-word sentences with intonations which are imitations of the slightly-exaggerated rising or falling contours used by adults in speaking to infants. Although these contours are non-contrastive, the intonation contour does function as a "container" for the utterance (see p. 89 above). As Weir's data shows, this period of non-contrastive imitative use of rising and falling contours lasts far longer than would appear to be the case to casual observers.

IV. Sometime after the second birthday, the infant begins to make contrastive use of intonations. In Miller and Ervin's study, this contrast was manifested as rising (question) versus falling (statement), but the authors' account of the acquisition of the rising contour hints at a bit of Skinnerian conditioning: "It may be that she learned the intonation by noting which sentences drew a response from the adult." (Miller and Ervin, 1964:29)

As noted above, Weir does not claim any contrastive status for the rising or falling or sustained contours used by her son Anthony. However, from her somewhat rare cases of pause-interrupted intonation contours, illustrated on p. 102 above, we infer the primitive beginnings of a syntactic intonation contrast which we would characterize in Hultzén's terms as low versus not-low phrase end, with the not-low contour expressing Bolinger's metaphor of "unfinished business." We could express this contrast in Lieberman's terms as the beginning of systematic use of the "marked" breath-group, but prefer not to do so, in view of the unfortunate physiological assumptions accompanying this terminology (see Chapter 2 above). It is interesting to note that the primitive nature of this contour contrast is matched in the stress aspects, which Weir describes separately, and summarizes in the following manner: "We can then say that the feature of stress as opposed to no stress has been well learned by the child, whereas the more complex contrastive use of various levels of stress has not." (30) If Bolinger is correct, the development of rising contours signalling "questioning" would be a later outgrowth from the more basic "unfinished business" intonation metaphor. Such a conception would be parallel to the transformational syntax view of yes-no questions as developing from disjunctive *S*'s of the *either-or* type. It should be noted that the experimental settings for the Miller-Ervin and Weir investigations were quite different, and that this may account for the differences in initial contrasts. Hopefully, the longitudinal studies of child language acquisition now in progress around the country will provide much more data on early syntactic contrasts in intonation.

V. There exists no firm data on the later development of intonation, and so we can only guess as to the probable chronology of further contrasts leading to the formation of the basic, or "core" adult system. We suspect that almost all of this enrichment takes place from the third through the eighth years. One of the earliest developments is probably a split of the not-low contour into two types of contour, one with a definite rise such as was seen in the *Were*[1] contours (Figs. 4.10-11, 4.13-14) indicating a question, the other having a slightly rising, sustained, or slightly falling phrase end, and expressing the non-questioning aspects of "unfinished business". Somewhat later, we suspect, the use of a fourth level (or, in contour terms, the use of an overhigh nuclear peak) becomes more sophisticated, and is used not only as part of emotional communication, but also as part of the syntactic system, expressing such things as rhetorical emphasis and contrastive stress. Later still, voice quality becomes usable for deliberate communication. That is, "angry" voice quality is manifested not only when one is actually angry, but also when one wishes to "act" angry. Also, the speaker begins to make systematic use of gradience in his intonation contours, and thereby begins to be able to suggest some of the nuances of communication detailed by Bolinger (1961a, 1964a). Through these ways, or through processes of a roughly similar nature, we believe that the speaker of English gradually builds his idiolectal variety of a "core" intonation system, common to all adult speakers of a broad geographical swath of American English (in the case of the speakers in our experiment, roughly "General American").

However, we have already stated our observation that some of our speakers (Speaker 12 in particular) displayed richer systems of intonational contrasts than others (see pp. 111-112 below), and we have already suggested the conclusion that these speakers continued to develop their intonation systems after they had finished the acquisition of virtually all of their syntactic and segmental phonology systems, so that in their language systems, intonation was both first learned and last learned.

Because the notion of differential competence among adult speakers of the same language flies in the face of recent doctrine concerning competence, we shall have to discuss such doctrine briefly.

Gleitman (1967) has contributed a searching criticism of those such as Katz who have extended Chomsky's (1965) notions concerning linguistic competence as a universal attribute of speakers into a kind of dogma, of which the following can be considered a mild manifestation:

> ...a necessary condition for something to be part of the subject matter of a linguistic theory is that each speaker be able to perform in that regard much as every other does. (Katz, 1964:415)

In clinging to this notion that competence must be universal and equal for all speakers, generative grammarians frequently ignore the great complexity which they themselves daily discover in the language, so that

> ...the mere production of a few, or a few thousand, dull but "previously unheard" sentences by the average speaker is taken as sufficient reason to endow him with all the wealth of the English language. Presumably the mere fact that a linguist's butcher can say "Good morning, Dr. Chomsky; the liver is fine today." serves as proof that in every butcher there have emerged the subtlest features of English syntactic structure. (Gleitman, 1967:37-38)

In order to test this assumption of universal and equal competence, Gleitman set up a series of experiments in which three groups of subjects were asked to supply paraphrases for a wide variety of compound nouns, some of which were relatively easy on both syntactic and semantic grounds (e.g. stone^2 bird^1 house^3), some semantically odd but grammatically possible (foot^1 house^3 bird^2), and some -- in Gleitman's terms -- "unfamiliar, bizarre, and ungrammatical" (eat^1 bird^3 house^2). In the first experiment, subjects were asked to freely generate paraphrases. In a second, much later experiment, they were asked to make a forced-choice of two possible paraphrases. The subjects were essentially monolingual English speakers from a mid-Atlantic or mid-Western background, but of three different kinds of educational background. Group A consisted of seven graduate students in various fields, Group B of seven undergraduates and college graduates who had no intention of doing graduate work, and Group C of eleven high-school graduates who had no intention of going to college (secretaries).

Gleitman found great differences in the performance of the differing groups, with no overlap at all in the number of errors for Group A and Group C. She concludes that it seems "a strategic error to assume that what is in Chomsky must therefore be in his butcher; and that what is in the grammar must therefore be in the mind of the user." (182) Furthermore, she generalizes her results by pointing out that studies of grosser aspects of linguistic organization (e.g. Miller and Isard, 1963; Johnson, 1965) have not shown significant individual differences in responses,

> However, when the response to more complex linguistic stimuli is examined (as here, and also cf. McNeill (1966) and Blumenthal (1966)) or when further skills and intuitions related to grammatical organization are tested (as here, the ability to paraphrase, and also cf. Maclay and Sleator (1960), the ability to classify sentences), individual differences in response-type begin to be found. And just as this more complex knowledge seems better developed in some individuals than in others, it seems to be better developed for certain fragments of the grammar than for others. (183)

Partly because of the persistence of the C-group subjects in repeating their errors systematically in the forced-choice situation (158), Gleitman felt confident that the differences which she found in performance directly reflected differences in linguistic competence. We accept her assessment of the data from those experiments. However, the tasks assigned subjects in our experiment were of a somewhat different nature, and the results seem to require a different, but not necessarily contrary, interpretation of the question of competence versus performance.

As we noted on pp. 28 and 42 above, we deliberately used our twelve speakers as listeners also, in order to determine whether they could correctly distinguish clearly contrasting test items produced by other speakers, even though they themselves had *not* produced a contrast on those same test items. The speakers gave overwhelming evidence that they could. For example, on the *Yes*$^{\text{angry}}$/*Yes*$^{\text{contained}}$ contrast, ten speakers performed well enough in producing the contrast to satisfy a high significance level, while two did not. But as listeners, the top ten and bottom two performed identically, with mean scores of 10.5 correct judgments out of a possible 12. On the most difficult test category, *Yes*^{3}(continuative)/*Yes*^{4}(repetitive question), six of the speakers performed at a significant level, and six did not. But the *bottom* six speakers performed somewhat better as listeners, with a mean score of 6.16 correct judgments, as compared with 4.83 for the top six speakers.

For our last example, let us look at the *Were*^{1}(question)/*Were*^{3} (subjunctive) contrast. As we noted on p. 61, Speaker 3 was the only one who seemed to lack such a contrast from the point of view of production (at least in the experimental situation). However, as a *listener*, Speaker 3 performed very well on this contrast. His score of 9 correct judgments almost matches the 9.27 mean score for the top 11 speakers. He had only one incorrect judgment, and two were neutralized. These neutralized judgments are very interesting. One of his neutralized judgments was of the test items produced by Speaker 7, and he was one of 24 who failed to hear a difference between those two stimuli. Speaker 3's other neutralized judgment was rendered upon *his own stimuli*! Thus, even though he knew (from the directions for that part of the listening test) what the contrast might be, he judged that he had failed, as a speaker, to produce such a contrast. (It is extremely probable that he recognized his own voice. In the first place, he had carefully monitored his own recording. Secondly, as we noted earlier, the two stimuli composing each test item were always preceded by a recording of the same speaker saying "Today is Monday." In the order in which he took the listening test, he had already heard himself saying that reference-tone phrase eight times, before he heard it again preceding his *Were*^{1}/*Were*^{3} stimuli.)

Speaker 3's performance as speaker and listener on the *Were*^{1}/*Were*^{3} contrast is only the most dramatic example of a phenomenon which occurred many times in our data. Such cases suggest that any attempt to deal with intonation contrasts in terms of a competence/performance scheme would have to include at least the following parameters:

Ability to produce a contrast clearly and easily, on a first try.	vs.	Ability to produce a clear contrast only after stumbling at first.
Ability to correct one's initial production error by oneself.	vs.	Ability to correct one's initial production error only after some "coaching". (N.B.: there was *no* "coaching" in our experiment.)
Ability to produce a clear contrast, and to hear it.	vs.	Ability to produce a clear contrast, but *not* to hear it. (The instances of this were so few and so isolated as to be attributable to momentary inattention or fatigue.)
Inability to produce a clear contrast and to hear it.	vs.	Inability to produce a clear contrast but ability to hear it.

Even if we eliminate production aided by coaching, on the grounds that such cases might possibly reflect imitation, rather than true generative ability, we are still left at least with the distinction between easy and hesitant production, and with the distinction between effective production and effective comprehension. The first distinction has generally been considered part of the realm assigned to performance (cf. Chomsky, 1965:4). But the second distinction raises some serious problems. First, an attempt to deal with differences in production and comprehension of speech might be interpreted as an attempt to resuscitate the notion of separate grammars-for-the-speaker and grammars-for-the-hearer (cf. Hockett, 1961), and would then be subject to the same objections in terms of needless duplication in the grammatical system (cf. Chomsky, 1964). Secondly, there remains the question of which of these aspects belong to competence and which to performance. Taking the second problem first, it would seem that the ability to perceive a meaningful difference between two speech samples must be taken as involving the use of underlying linguistic competence, particularly when the speech samples in question are neither bizarre nor idiosyncratic, and when the perceptual process makes use of no non-linguistic information.

But what about the production of intonation contrasts? We believe that this also must be attributed to underlying competence (as realized through performance mechanisms). Such a decision does not necessitate setting up separate grammars for speaking and hearing. Instead, discrepancies between production and comprehension behavior might be explained by means of a scheme for linguistic competence arranged on a continuum from "active" to "latent". In such a scheme, production would demand that the item to be produced would have to lie toward the "active" end of the continuum, unless the speaker is to refresh his memory by resorting to a written grammar, thesaurus, phonetics manual, or language teacher. Comprehension, on the other hand, would function quite well with items anywhere on the scale, including the "latent" end.

According to this view, Speaker 3 would be said to have the *Were*^{1}/ *Were*^{3} contrast within his linguistic competence, but on the "latent" side of the continuum. Theoretically, it would take only a small amount of speech training to teach him to produce such a contrast.

To take a more delicate case, Speaker 12, on her first recording of the dialogue, said "If they were black..." instead of "Were they black..." with a subjunctive intonation for *Were*^{3}. While monitoring her recording, she caught this deviation from the script with no aid from the experimenter, and immediately recorded a subjunctive intonation for "were they black" which, when contrasted with her *Were*^{1}, received a very good C-score of 80 (from 45 correct listener judgments, and 5 incorrect). We would interpret this case as indicating that Speaker 12 had both of these alternate means of expressing the subjunctive in the active portion of her competence, but with the "if..." subjunctive more active than the purely intonational subjunctive.

Do our notions of linguistic competence and our data run counter to the conclusions (p. 104 above) of Gleitman? We do not believe so. Although the stimuli and the experimental conditions of our experiments were quite different, Gleitman's use of free paraphrase versus forced-choice of given paraphrases would seem to correspond roughly with our conception of active versus latent competence, and her denial of Katz's assumption of universal and equal competence was based on the dismal performance of her C-group subjects on *both* kinds of tasks. Also of interest are the cases of our two "worst" speakers in the *Yes*^{3}/*Yes*^{4} test category. Speaker 4's performance as a speaker yielded a C-score of -10, while Speaker 5 (Fig. 4.9) received a -30. As listeners, they were also at the bottom of the group on this category. Speaker 4 made four correct judgments, seven incorrect, with one neutralized. She also heard and correctly categorized the two deliberately neutralized test items included in the listening test. Speaker 5 made only three correct judgments, seven incorrect, and two neutralized. While he heard the deliberately neutralized test items as neutralized, he labelled them wrongly (i.e. as BB instead of AA). There was also a striking qualitative difference in their performance as listeners. Speaker 4's four correct judgments were of stimuli produced by those speakers who ranked first, second, third, and fifth in their speaking performance on this category. These four speakers had a mean C-score on this category of 59.5, well above the 38.60 needed for a significance level of <.01. In contrast, Speaker 5's three correct judgments were of speakers whose mean C-score on this category was only 31.33 (*below* the significance level). These qualitative differences suggest the following as a possible interpretation: Speaker 4's inability to produce the *Yes*^{3}/*Yes*^{4} contrast and inability to hear and correctly categorize the contrast, except when it was produced in a very clear manner, indicate that her competence on this contrast is limited to the extreme latent end of an active/latent competence continuum. On the other hand, Speaker 5's somewhat anomalous performance as a speaker, and his poor, haphazard performance as a listener suggest that he lacks competence on this contrast or, alternatively, that his scheme for *Yes*^{3} and *Yes*^{4} has a very poor "fit" with the competence scheme of the other speakers.

These two instances do not, of course, prove anything. They do, however, demonstrate that the notion of latent competence need not be a garbage dump for unsolved problems, and they at least suggest the possibility of finding a genuine lack of competence among adult speaker/hearers in areas of the grammar additional to those studied by Gleitman (cf. also p. 112 below). This possibility deserves further study.

Differences in Effectiveness among Speakers

As we noted earlier (p. 28 above), in planning our experiment we were interested not only in the kinds of intonational gestures used to signal intonation contrasts, but also in the possibility that some speakers would be more effective than others in communicating such contrasts. We therefore used two groups of speakers. The "sophisticated" group consisted of six graduate students, subdivided into a male and female from the fields of linguistics, English literature, and speech (including oral interpretation). This group was matched with six "naive" speakers (three male, three female) of the same mean age who were undergraduates with no advanced work in foreign languages or linguistics, literature, or speech. We expected that the "sophisticated" speakers would be more effective in communicating syntactic and emotional information than the "naive" speakers, and we expected especially good performance from the graduate students in English and speech, for reasons which we discussed on pp. 40-41.

The results for overall speaker performance, summarized in Table III (p. 108), were rather different from what we had expected. The primary cause for this was the fact that *all* the speakers communicated effectively; the worst individual speaker mean C-scores were in the low sixties,

Table III: Overall Speaker Performance

	Y^1/Y^4	Y^1/Y^2	Y^3/Y^4	W^1/W^3	W^1/W^2	W^2/W^3	$\frac{Y^{un.}}{Y^{emph.}}$	$\frac{R^{un.}}{R^{emph.}}$	$\frac{\text{belief}}{\text{disb.}}$	$\frac{\text{bored}}{\text{int.}}$	$\frac{\text{agree}}{\text{disag.}}$	$\frac{\text{calm}}{\text{angry}}$	$\frac{\text{angry}}{\text{cont.}}$	Mean C-score
Sp. 1	92	84	26	80	-10	72	92	94	86	100	94	94	82	75.85
Sp. 2	90	96	-4	94	52	52	86	100	80	-62	52	90	84	62.30
Sp. 3	90	80	36	-6	44	-28	100	96	46	96	90	98	84	63.54
Sp. 4	96	74	-10	82	58	68	98	86	94	78	18	96	88	71.23
Sp. 5	76	94	-30	84	76	82	82	94	62	84	88	96	32	70.77
Sp. 6	62	70	0	92	48	60	100	100	70	98	98	96	28	70.92
Sp. 7	92	86	42	44	64	4	78	86	90	92	92	100	100	74.61
Sp. 8	44	92	-6	64	74	72	100	90	44	100	100	98	80	73.23
Sp. 9	92	96	70	90	74	32	96	84	56	96	58	100	76	78.46
Sp. 10	52	88	62	72	38	48	90	96	42	100	-84	94	100	61.38
Sp. 11	40	82	26	20	8	28	92	98	40	96	88	100	94	62.46
Sp. 12	96	78	70	80	70	36	100	86	92	100	86	98	98	83.85

far above the figure of 10.31 (N=13 categories x 50 listeners = 650) needed for a .01 significance level on a two-tailed test. Even those test categories deliberately included because of their supposed difficulty yielded effective communication on the part of almost all the speakers, with the exception of the very difficult *Yes*^{3}/*Yes*^{4} category. This high level of overall performance prevented the emergence of any gross differences in overall performance between the two groups, as indicated by the following figures:

Table IV

	Mean C-score (all 13 categories)	Mean C-score (best 12 categories)
Speakers 1-6 "naive"	69.10	76.80
Speakers 7-12 "sophisticated"	72.33	78.49

However, when we reconsider Gleitman's observation (p. 106 above) that performance differences which appear non-existent in tasks relating to grosser aspects of language become evident in the case of more complex linguistic entities, and when we contrast the performance of our two groups of speakers in "easy" versus "hard" categories, then subtle but significant differences begin to emerge:

Table V

Mean C-scores

	Two "easy" categories		Two "hard" categories	
	Yes$^{\mathrm{un}}$/*Yes*$^{\mathrm{emph}}$	*Yes*$^{\mathrm{calm}}$/*Yes*$^{\mathrm{angry}}$	*Yes*^{3}/*Yes*^{4}	*Yes*$^{\mathrm{angry}}$/*Yes*$^{\mathrm{contained}}$
Sp. 1-6	93.0	95.0	3.0	66.33
Sp. 7-12	90.0	98.33	44.0	91.33

As the above table shows, the high scores for all speakers in the "easy" categories prevented differences of any significance at all from emerging. But the "hard" categories present a quite different picture. A Mann-Whitney *U*-test (Winer, 1962:623) shows the difference between the "naive" and "sophisticated" groups to be significant at a level of .025 for *Yes*^{3}/*Yes*^{4}, and .066 for *Yes*$^{\mathrm{angry}}$/*Yes*$^{\mathrm{contained}}$ (the use of a *t*-test would have yielded even higher significance levels). These results appear to support Gleitman's contention that equal competence in the simpler or more familiar parts of the grammar does not necessarily extend to more complex or esoteric parts of the grammar.

Some Comments on Individual Speakers

Speaker 12, a female graduate student in speech whose specialty was oral interpretation, stood above the other speakers. Her overall mean C-score of 83.85 was more than five points above that of the second-ranking performer, Speaker 9 (a male graduate student in English), and was more than 14 points above the mean C-score for Speakers 1-11. We shall discuss her performance in greater detail in the next section.

Speakers 7 and 8, the graduate students in linguistics, did better than we had anticipated. Speaker 7's effectiveness resulted from his willingness to use gross effects (including occasional falsetto voice). His stimuli sometimes produced snickers among the listeners, but succeeded in communicating the desired contrasts. Speaker 8 was simply a good performer, using an effective mixture of contour shape and voice quality cues to signal the contrasts. Speaker 1 did the same.

Both Speaker 10 and Speaker 11 failed to communicate as effectively as they might have because of an excessive reliance upon subtle voice quality cues, and occasional choices of anomalous contour shapes. This was especially the case with Speaker 11, the male graduate student in speech who, because of his work in oral interpretation, should have done better. Although his anomalous contours might have reflected the influence of his Norwegian grandfather (see. p. 41 above), we suspect that they resulted more from a "manly" early-Marlon Brando acting style. Speaker 9 also affected a "manly" tone, with lowered fundamental frequency, falling contours ending in creaky voice, and a slight "biting" edge to the voice similar to that employed by Speaker 11. However, Speaker 9 did not make Speaker 11's mistake of excessively flattening most of his f_0 contours. In fact, Speaker 9 sometimes used very lively contours.

A "Sophisticated" System

On pp. 100-102 we sketched a possible chronology leading to the development of a "core" intonation system, one which would be common to all the speakers in this study, and to the type of General American dialect they employ. Even the development of that "core" system entails, in our opinion, the acquisition and/or refinement of intonation features long after this process has generally been supposed to be completed.

Speaker 12 provides further evidence that intonation features exist, and are clearly perceived by listeners, which are unlikely to be at the productive command of an eight-year-old, let alone a three-year-old. In achieving her extremely high overall mean C-score of 83.85, Speaker 12 used the following arsenal of contour shapes (combined with amplitude variations and with a subtle but expressive repertoire of paralintuistic cues):

Yes^{1}:	scooped nucleus, low terminal rise
Yes^{2}:	very short, very level
Yes^{3}:	scooped nucleus, moderate terminal rise
Yes^{4}:	very high rise
Were^{1}:	scooped nucleus, moderate terminal rise
Were^{2}:	scooped nucleus, low terminal rise
Were^{3}:	high nuclear rise, part-fall, level end
Yes$^{\text{un}}$:	relatively short, level
Yes$^{\text{emph}}$:	long, moderate rise, fall, level end
Ridic$^{\text{un}}$:	slight early rise, level
Ridic$^{\text{emph}}$:	very high nuclear rise, fall
Yes$^{\text{belief}}$:	low fall, level end
Yes$^{\text{disbel.}}$:	early fall, long level end
Yes$^{\text{bored}}$:	early part-fall, long level end
Yes$^{\text{int.}}$:	slight rise, slight nuclear scoop, high terminal rise, level end
Yes$^{\text{agree}}$:	nuclear rise, part-fall, level end
Yes$^{\text{disag.}}$:	long level nucleus, abrupt part-fall
Yes$^{\text{calm}}$:	part-fall, level end
Yes$^{\text{angry}}$:	high nuclear rise, full fall
Yes$^{\text{cont.}}$:	very high nuclear rise, full fall, level end

How does such a varied system develop? We believe that the main characteristics of the "sophisticated" system of Speaker 12 are the following: (1) the skilled use of scooped nuclear segments (essential for *Yes*^{3}, helpful elsewhere); (2) a fine degree of gradience, particularly at the end of contours, so that, for example, part-falls are distinctly differentiated from full-falls, and their accompanying meaning differences can be systematically employed; (3) the skilled integration of paralinguistic cues (particularly voice quality) with contour shape to communicate emotional attitudes. As with the gradience phenomena, this represents a refinement of elements already existing in the simpler system used by the adult population in general. Thus, the infant unconsciously displays "angry" voice quality, the child learns how to "act" angry with his voice, the typical "naive" adult speaker effectively communicates "contained anger", as well as some other emotional attitudes, while the "sophisticated" adult speaker achieves near-perfect communication of "contained anger" and of a wide variety of other emotional states; (4) the use of extended vocal range: the best communicators not only made subtle distinctions between certain kinds of contours, but also extended the range of, say, the rise on *Yes*^{4}, in order to clarify the contrast with the moderate rise of *Yes*^{3}. Speaker 12's frequency range might be analyzed as having not only an "overhigh", but an "over-overhigh", featuring a deliberate voice break. Speaker 9 seemed to manifest two varieties of full-fall. The second, more "decisive" variety dropped lower than the first, right into creaky voice.

Taming the Servant: Should We Teach Intonation?

We believe that the extremely effective performance of Speaker 12 in communicating syntactic and emotional information by means of intonation resulted from her extensive experience as a student, part-time teacher, and performer in the field of oral interpretation. We do not regard the performance of Speaker 11, who had a similar background, as constituting a piece of negative evidence. As we noted earlier, the mediocre performance of Speaker 11 and Speaker 10 resulted largely from their tendency to use ambiguous f_o contours, and to depend upon paralinguistic cues to signal the intended meaning. The particular experimental situation here tended to minimize the effectiveness of such a strategy, since the usual carrier phrase consisted only of the word *yes*. If the carrier phrase for *Yes*$^{\text{belief}}$/*Yes*$^{\text{disbelief}}$ (see p. 76 above) had been *Yes, I heard it*, we are certain that Speaker 11 would have had a better opportunity to communicate his disbelief to the listeners. Similarly, if the carrier phrase for *Yes*$^{\text{agreeable}}$/*Yes*$^{\text{disagreeable}}$ had been something like *Yes, I did it*, Speaker 10 (see pp. 82-85 above) would probably have been able to communicate her attitudes better, and would not have received a C-score of -84. But perhaps this kindness is out of place. Our choice of brief carrier phrases was deliberate and, we believe, well motivated (see pp. 31-32 above). It may be that the cases of Speaker 12 and Speaker 11 simply demonstrate that in every field there are good and bad students. Yet this observation does not end the argument, because it is possible that better teaching would produce better students.

Although many well-known studies of the intonation of British English (e.g. the works of Armstrong and Ward, Palmer, Kingdon, and Schubiger) have derived from an interest in teaching English as a second language, and although at least one outstanding American treatment (Pike, 1945) had a similar origin, the notion of teaching intonation to native speakers of English must initially seem strange. After all, native speakers of English know the difference between "normal" and "repetitive" questions, and are not likely to annoy listeners by making universal use of the "repetitive" type. Nor are they likely to horrify social gatherings by asking *When are we going to eat Susie?* instead of *When are we going to eat, Susie?*

However, recent attempts to teach standard American English to speakers of non-standard dialects have broadened the scope of English as a second language. To take a delicate case, most authorities would now agree that the 18-year-old student who tries to add subjuncitve *were* to his idiolect faces problems akin to those in learning a construction in a foreign language. Our data add that it is not enough to learn when to say *were*. One must also learn the appropriate intonation.

The kind of intonation teaching we have in mind would be quite different from the kind of speech instruction which arose from the teaching of "elocution" in 19th century America, and flourished well into the present century. A last gasp of that movement was represented by the aging spinster who visited our high school English class for a single week during the junior and senior years, and whose *entire* curriculum consisted of having us memorize and give m-e-l-l-i-f-l-u-o-u-s individual reading of a dreadful poem which began,

It is not what you say,
But how you say it...

Teachers of intonation should have a thorough background in speech, linguistics, and literary analysis. They should be aware that intonation signals both syntactic and emotional meaning, that this communication consists of several parameters, not just voice quality, and that voice quality is often an inadequate means of signalling even emotional meaning. Lastly, they should be able to demonstrate to their students how the study of intonation can enrich their understanding and enjoyment of literature.

But is all this really necessary? Doesn't context always supply information hidden or confused by inadequate intonation? Does communication ever hang upon brief phrases with little or no context? We could reply by reminding the reader that the extremely sloppy uses of "ironic" intonation by flight controllers and crew during the recent Apollo flights depended for their success upon the fact that the men involved had spent hundreds of hours listening to each others' voices. In flights involving international personnel, communication of meaning through intonation would have to be improved at the risk of, if not aborted flights, at least ruffled tempers in space.

However, we prefer to close with an example somewhat closer to home. In their interesting micro-analysis of the first five minutes of an interview of a new patient by a psychiatrist, Pittenger, Hockett, and Danehy (1960) detail a major breakdown in emotional communication near the beginning of the interview. The authors observe that the therapist "perhaps believes that at certain points in an interview the patient needs to be told that he is being heard and understood, and that he should continue; but that this information should be conveyed in a way which is free of any overtones of emotional reaction or moral judgment-passing on the therapist's part -- except, possibly, for a very generalized sympathy." (27b) Unfortunately, in his attempt to achieve such an "opaque" intonation, the therapist uttered the phrase *Yeah*, with an intonation transcribed by the authors as /^{2}yéh^2#/. The authors explain the patient's distressed reaction on the grounds that the therapist's utterance resembled "a drawled ^{3}yéah^2#", which they say would mean something like "How often I've heard things just like this! Here we go again!" (29b) Our interpretation would differ slightly, since our data show that very flat contours indicate *more* unpleasantness than the falling contour cited by the authors. But the main point here is that the proper way of indicating this interest in hearing more, together with a "generalized sympathy", would be through the use of a contour resembling the better examples of *Yes*^{3} (see pp. 57-60). Furthermore, the therapist should have sensed that a lively contour is far more effective in communicating "interest" than a flattened one (see p. 79 above).

If a highly-trained psychiatrist can make such an inappropriate choice of intonation, how many minor tragedies of communication occur in America every day?

BIBLIOGRAPHY

Abe, I. (1955), "Intonational Patterns of English and Japanese," *Word, 11,* 386-398.

Abe, I. (1957-58), "On Japanese Intonation: An Experiment," *Lingua, 7,* 183-194.

Abercrombie, D. (1967), *Elements of General Phonetics,* Chicago: Aldine.

Armstrong, L.E. and I.C. Ward (1926), *Handbook of English Intonation,* Leipzig and Berlin: Teubner.

Artemov, V.A. (1969), "Speech Intonation," in *Study of Sounds, 14,* 1-19.

Atkinson, K. (1968), "Language Identification from Non-Segmental Cues," *Working Papers in Phonetics, 10* (UCLA Phonetics Laboratory), 85-89.

Bever, T., Fodor, J.A. and W. Weksel (1965), "On the Acquisition of Syntax: A Critique of 'contextual generalization'," *Pscyhol. Rev., 72,* 467-482.

Bierwisch, M. (1966), "Regeln für die Intonation dutscher Sätze," *Studia Grammatica, 7,* 99-201.

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

Bolinger, D.L. (1949), "Intonation and Analysis," *Word, 5,* 248-254.

_____(1951), "Intonation: Levels vs. Configurations," *Word, 7,* 199-210.

_____(1957), *Interrogative Structures of American English (The Direct Question),* American Dialect Society publication no. 28, University, Ala.: University of Alabama Press.

_____(1957-58a), "On Intensity as a Qualitative Improvement of Pitch Accent," *Lingua, 7,* 175-182.

_____(1957-58b), "Intonation and Grammar," *Language Learning, 8,* 31-38.

_____(1958), "A Theory of Pitch Accent in English," *Word, 14,* 109-149.

_____(1960), "Linguistic Science and Linguistic Engineering," *Word, 16,* 374-391.

_____(1961a), *Generality, Gradience, and the All-or-None,* The Hague: Mouton.

_____(1961b), "Ambiguities in Pitch Accent, *Word, 17,* 309-317.

_____(1961c), "Contrastive Accent and Contrastive Stress, *Language, 37,* 83-96.

_____(1964a), "Intonation as a Universal," *Proc. 9th Int'l. Cong. Ling.,* The Hague: Mouton, 832-848.

_____(1964b),"Around the Edge of Language: Intonation," *Harvard Educational Review, 34,* 282-296.

_____(1965), *Forms of English: Accent, Morpheme, Order,* ed., I. Abe and T. Kanekiyo, Cambridge, Mass.: Harvard Univ. Press.

_____(1968), *Aspects of Language,* New York: Harcourt, Brace, and World.

_____ and L.J. Gerstman (1957), "Disjuncture as a Cue to Constructs," *Word, 13,* 246-255.

Bosma, J.F., Lind, J., and H.M. Truby (1964), "Respiratory Motion Patterns of the Newborn Infant in Cry," in *Physical Diagnosis of the Newly Born, Report of the Forty-Sixth Ross Conference on Pediatric Research,* J.L. Kay, ed., Ross Laboratories, Columbus, Ohio, 103-111.

_____, Truby, H.M., and J. Lind (1965), "Cry Motions of the Newborn Infant," in *Newborn Infant Cry,* ed. J. Lind, *Acta Paediatrica Scandinavia, Suppl. 163,* Uppsala, 61-92.

Braine, M.D.S. (1963a), "The Ontogeny of English Phrase Structure: The First Phase," *Language, 39,* 1-13.

_____(1963b), "On Learning the Grammatical Order of Words," *Psychol. Rev., 70,* 323-348.

_____(1968, forthcoming), "The Acquisition of Language in Infant and Child," to appear in C. Reed (ed.), *The Learning of Language.*

Bronstein, A.J. (1960), *The Pronunciation of English: An Introduction to Phonetics,* New York: Appleton-Century-Crofts.

Brooks, N. (1960), *Language and Language Learning,* New York: Harcourt, Brace and World.

Carroll, J.B. (1961), "Language Development in Children," in S. Saporta (ed.), *Psycholinguistics: A Book of Readings,* New York: Holt, 331-345.

Chomsky, N. (1964a), *Current Issues in Linguistic Theory,* The Hague: Mouton.

_____(1964b), Discussion of W. Miller and S.M. Ervin, "The Development of Grammar in Child Language," in Bellugi and Brown, *The Acquisition of Language*, 34-38.

_____(1965), *Aspects of the Theory of Syntax*, Cambridge, Mass.: MIT Press.

_____(1967), "The Formal Nature of Language," appendix to E.H. Lenneberg, *Biological Foundations of Language*, New York: Wiley.

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

Cowan, M. (1936), "Pitch and Intensity Characteristics of Stage Speech," *Archives of Speech, Suppl. 1*, 1-92.

Daneš, F. (1960), "Sentence Intonation from a Functional Point of View," *Word, 16*, 34-54.

DeLattre, Pierre (1969), "Syntax and Intonation: A Study in Disagreement," in *Study of Sounds, 14*, Tokyo: Phonetic Society of Japan, 21-40.

Denes, P. (1959), "A Preliminary Investigation of Certain Aspects of Intonation," *Language and Speech, 2*, 106-122.

_____(1965), "On the Motor Theory of Speech Perception," *Proc. 5th Int'l. Cong. Phon. Sci.*, Basel, 252-258.

_____, and J. Milton-Williams (1962), "Further Studies in Intonation," *Language and Speech 5*, 1-14.

Ervin, S.M. and W.R. Miller (1963), "Language Development," in *Child Psychology* Sixty-second Yearbook, Part 1, National Society for the Study of Education), Chicago: Univ. of Chicago Press, 108-143.

Ervin-Tripp, S.M. (1966), "Language Development," in L.W. Hoffman and M.L. Hoffman (ed.), *Review of Child Development Research, 2*, New York: Russell Sage Foundation, 55-105.

Flanagan, J.L. (1955), "A Difference Limen for Vowel Formant Frequency," *J. Acoust. Soc. Am., 27*, 613-617.

_____(1957), "Estimates of the Maximum Precision Necessary in Quantizing Certain 'Dimensions' of Vowel Sounds," *J. Acoust. Soc. Am., 29*, 533-534.

_____(1965), *Speech Analysis Synthesis and Perception*, New York: Academic Press.

_____(1968), "Studies of a Vocal-Cord Model Using an Interactive Laboratory Computer," *Preprints for the Kyoto Speech Symposium*, C-1-o--C-1-6.

Fodor, J.A. and T.G. Bever (1965), "The Psychological Reality of Linguistic Segments," *J. Verb. Learn. and Verb. Behav.*, *4*, 414-420.

Fraser, C., Bellugi, U., and R. Brown (1963), "Control of Grammar in Imitation, Comprehension, and Production," *J. Verb. Learn. and Verb. Behav.*, *2*, 121-135.

Fry, D.B. (1955), "Duration and Intensity as Physical Correlates of Linguistic Stress," *J. Acoust. Soc. Am.*, *35*, 765-769.

_____(1958), "Experiments in the Perception of Stress," *Language and Speech*, *1*, 126-152.

_____(1960), "Linguistic Theory and Experimental Research," *TPS*, 13-39.

_____(1966), "The Development of the Phonological System in the Normal and the Deaf Child," in F.L. Smith and G.A. Miller (eds.), *The Genesis of Language*, 187-206.

Galunov, V.I. and L.A. Chistovich (1965), "Relationship of Motor Theory to the General Problem of Speech Recognition," *Soviet Physics-Acoustics*, *11*, 357-365. Originally appeared in *Akustichekii Zhurnal*, *11*, 417-426.

Gårding, E. and A.S. Abramson (1965), "A Study of the Perception of some American English Intonation Contours," *Studia Linguistica*, *19*, 61-79.

Gårding, E. and L.J. Gerstman (1960), "The Effect of Changes in the Location of an Intonation Peak on Sentence Stress," *Studia Linguistica*, *14*, 57-59.

Gleitman, L.R. (1967), *Compound Nouns and English Speakers*, Philadelphia, Penna.: Eastern Penna. Psychiatric Institute.

Grégoire, A. (1937), *L'Apprentissage du language*, Liège, Belgium: l'Université de Liège.

Hadding-Koch, K. (1956), "Recent Work on Intonation," *Studia Linguistica*, *10*, 77-96.

_____(1961), *Acoustico-phonetic Studies in the Intonation of Southern Swedish*, Lund: Gleerup.

_____(1964), Review of A.V. Isačenko and H-J. Schädlich (1964), *Untersuchungen über die deutsche Satzintonation*, *Studia Linguistica*, *18*, 122-131.

_____ and M. Studdert-Kennedy (1964), "An Experimental Study of Some Intonation Contours," *Phonetica, 11,* 175-185.

Halle, M. (1962), "Phonology in Generative Grammar," *Word, 18,* 54-72.

_____ and K.N. Stevens (1964), "Speech Recognition: A Model and a Program for Research," in J.A. Fodor and J.J. Katz (eds.), *The Structure of Language,* 604-612.

Halliday, M.A.K. (1963a), "The Tones of English," *Archivum Linguisticum, 15,* 1-28.

_____(1963b), "Intonation in English Grammar," *TPS*, 143-169.

_____(1967), *Intonation and Grammar in British English,* The Hague: Mouton.

Harris, Z. (1944), "Simultaneous Components in Phonology," *Language, 20,* 181-205.

Helmholtz, H.L.F. (1954), *On the Sensations of Tone* (trans. A.J. Ellis), New York, Dover. Originally published as *Die Tonempfindung (1863),* Berlin.

Hockett, C.F. (1955), *A Manual of Phonology (IJAL Memoir 11),* Baltimore.

_____(1961), "Grammar for the Hearer," in R. Jakobson (ed.), *Structure of Language and its Mathematical Aspects (Proceedings of Symposia in Applied Mathematics, 12),* Providence, R.I.: American Mathematical Society, 220-236.

Hultzén, L.S. (1955), "Stress and Intonation," *General Linguistics, 1,* 35-43.

_____(1956), "'The Poet Burns' Again," *American Speech, 31,* 195-201.

_____(1957), "Communication in Intonation: General American," in *Study of Sounds, 2,* Tokyo: Phonetic Society of Japan, 317-333.

_____(1959), "Information Points in Intonation," *Phonetica, 4,* 107-120.

_____(1962), "Significant and Nonsignificant in Intonation," *Proc. 4th Int'l. Cong. Phon. Sci.,* The Hague: Mouton, 658-661.

_____(1964), "Grammatical Intonation," in *In Honour of Daniel Jones,* D. Abercrombie *et. al.* (eds.), London: Longmans, Green, 85-95.

Hunt, K. (1965), *Grammatical Structures Written at Three Grade Levels,* Champaign, Ill.: NCTE.

Isačenko, A.V. and H-J. Schädlich (1963), "Erzeugung Künstlicher deutscher Satzintonationen mit zwei kontrastierenden Tonstufen," *Monatsber. deut. Akad. Wiss. 5*, Berlin, 365-372.

Jakobson, R. (1941), *Kindersprache, Aphasie, und allgemeine Lautgesetze*, Uppsala. English version: (1968), *Child Language, Aphasia, and Phonological Universals*, The Hague: Mouton.

_____ and M. Halle (1956), *Fundamentals of Language*, The Hague: Mouton.

Jassem, W. (1952), *Intonation of Conversational English*, Wroclaw, Poland.

Jespersen, O. (1922), *Language: Its Nature, Development and Origin*, London: Allen and Unwin.

Johnson, Neil F. (1965), "The Psychological Reality of Phrase-Structure Rules," *J. Verb. Learn. and Verb. Behav., 4*, 469-475.

Jones, D. (1909), *Intonation Curves*, Leipzig and Berlin: Teubner.

_____(1956), *The Pronunciation of English (4th ed.)*, Cambridge: Cambridge Univ. Press.

Joos, M. (1948), *Acoustic Phonetics (Language*, Monograph No. 23, Suppl. to *Language, 24.2)*.

_____(1964), "Language and the School Child," *Harvard Educational Review, 34*, 203-210.

Juhasz, F. (1963), Review of L. Elekfi (1962), *Vizsgálatok a hanglejtés megfigyelésének módjaihoz (Experiments on the Observation of Intonation), Word, 19*, 122-126.

Katz, J.J. (1964), "Semi-Sentences," in J.A. Fodor and J.J. Katz, *The Structure of Language*, Englewood Cliffs, N.J.: Prentice-Hall, 400-416.

Kim, C-W. (1968), Review of P. Lieberman (1967), *Intonation, Perception, and Language*, *Language, 44*, 830-42.

Kingdon, Roger (1958), *The Groundwork of English Intonation*, London: Longmans.

Kozhevnikov, V.A. and L.A. Chistovich *et. al.* (1965), *Speech: Articulation and Perception*, Moscow and Leningrad. Translation JPRS 30.543. U.S. Department of Commerce.

Kurath, H. (1964), *A Phonology and Prosody of Modern English*, Ann Arbor: Univ. of Mich. Press.

Ladefoged, P. (1958), "Syllables and Stress," *Misc. Phonetica, 3,* 1-14.

_____(1962a) "Sub-Glottal Activity during Speech," *Proc. 4th Int'l. Cong. Phon. Sci.,* The Hague: Mouton.

_____(1962b), *The Nature of Vowel Quality,* Coimbra: Laboratoria da Phon. Exp.

_____(1967), *Linguistic Phonetics (Working Papers in Phonetics, 6),* Los Angeles: UCLA Phonetics Laboratory.

_____(1968), "Linguistic Aspects of Respiratory Phenomena," in *Sound Production in Man* (conference sponsored by N.Y. Academy of Sciences), New York.

_____ and D.E. Broadbent (1957), "Information Conveyed by Vowels," *J. Acoust. Soc. Am., 29,* 98-104.

_____ and V. Fromkin (1968), "Experiments on Competence and Performance," *IEEE Transactions on Audio and Electroacoustics, AU-16.1,* 130-136.

_____ and N.P. McKinney (1963), "Loudness, Sound Pressure, and Subglottal Pressure in Speech," *J. Acoust. Soc. Am., 35,* 454-460.

Lane, H.L. (1965), "The Motor Theory of Speech Perception: A Critical Review," *Psychol. Rev., 72,* 275-309.

_____(1967), "A Behavioral Basis for the Polarity Principle in Linguistics", *Language, 43,* 494-511.

Lee, W.R. (1956), "English Intonation: A New Approach," *Lingua, 4,* 345-371.

Lees, R.B. (1960), Review of D.L. Bolinger (1957), *Interrogative Structures of American English, Word, 15,* 119-125.

Lehiste, I. and G.E. Peterson (1961), "Some Basic Considerations in the Analysis of Intonation," *J. Acoust. Soc. Am., 33,* 419-425.

Lenneberg, E.H. (1964), "A Biological Perspective of Language," in E.H. Lenneberg (ed.), *New Directions in the Study of Language,* Cambridge: MIT Press.

_____(1967), *Biological Foundations of Language,* New York: Wiley.

Leopold, W.F. (1939), *Speech Development of a Bilingual Child,* Evanston: Northwestern Univ. Press.

Lewis, M.M. (1951), *Infant Speech,* New York: Humanities Press.

_____(1963), *Language, Thought and Personality in Infancy and Childhood,* New York: Basic Books.

Liberman, A.M. (1957), "Some Results of Research on Speech Perception," *J. Acoust. Soc. Am., 29,* 117-123.

_____, Cooper, F.S., Harris, K.S., and P.F. MacNeilage (1963), "A Motor Theory of Speech Perception," *Proc. of the Speech Communication Seminar,* Speech Transmission Laboratory, RIT, Stockholm.

Lieberman, P. (1960), "Some Acoustic Correlates of Word Stress in American English," *J. Acoust. Soc. Am., 32,* 451-454.

_____(1963), "Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech," *Language and Speech, 6,* 172.

_____(1965), "On the Acoustic Basis of the Perception of Intonation by Linguists," *Word, 21,* 40-54.

_____(1967), *Intonation, Perception, and Language* (Research Monograph *38*), Cambridge: MIT Press.

_____(1968), "Phonologic Features and Muscular Commands, with Particular Reference to the Larynx," *Preprints for the Kyoto Speech Symposium,* C-4-1--C-4-5.

_____, Knudson, R., and J. Mead (1969, forthcoming), "Determination of the Rate of Change of Fundamental Frequency with Respect to Subglottal Air Pressure during Sustained Phonation," *J. Acoust. Soc. Am., 41.*

Lindblom, B. (1962), "Accuracy and Limitations of Sona-Graph Measurements," *Proc. 4th Int'l. Cong. Phon. Sci.,* The Hague, Mouton.

Luria, A.R. and F.I. Yudovich (1966), *Speech and the Development of Mental Processes in the Child* (ed. Joan Simon), London: Staples Press.

McCarthy, D. (1954), "Language Development in Children," in L. Carmichael (ed.), *A Manual of Child Psychology (*2nd ed.), New York: Wiley, 492-630.

Magdics, K. (1963), "Research on Intonation during the Past Ten Years," *Acta Linguistica, 13,* 133-165.

Menyuk. P. (1963), "A Preliminary Evaluation of Grammatical Capacity in Children," *J. Verb. Learn. and Verb. Behav., 2,* 429-439.

Meyer-Eppler, W. (1957), "Realization of Prosodic Features in Whispered Speech," *J. Acoust. Soc. Am., 29,* 104-106.

Miller, G.A. (1956), "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychol. Rev., 63,* 81-96.

_____(1962), "Some Psychological Studies of Grammar," *American Psychologist, 17,* 748-762.

_____(1964), "Language and Psychology," in E.H. Lenneberg (ed.), *New Directions in the Study of Language,* Cambridge: MIT Press.

_____ and S. Isard (1963), "Some Perceptual Consequences of Linguistic Rules," *J. Verb. Learn. and Verb. Behav., 2,* 217-228.

Miller, W. and S.M. Ervin (1964), "The Development of Grammar in Child Language," in U. Bellugi and R. Brown (eds.), *The Acquisition of Language (Monographs of the Society for Research in Child Development, 29.1),* 9-34.

O'Connell, D.C., Turner, E.A., and L.A. Onuska (1968), "Intonation, Grammatical Structure, and Contextual Association in Immediate Recall," *J. Verb. Learn. and Verb. Behav., 7,* 110-116.

O'Donnell, R.C., Griffin, W.J., and R.C. Norris (1967), *Syntax of Kindergarten and Elementary School Children,* Champaign, Illinois: NCTE.

Ohala, J. (1969), *Aspects of the Control and Production of Speech,* UCLA Dissertation. To appear as *Working Papers in Phonetics, 14, 1969.*

_____ and M. Hirano (1967), "Studies of Pitch Change in Speech," in *Working Papers in Phonetics, 7,* UCLA Phonetics Laboratory, 80-84.

Öhmann, S. and J. Lindqvist (1966), "Analysis-by-Synthesis of Prosodic Pitch Contours," *Quarterly Progress and Status Report, 4,* Stockholm: Speech Transmissions Laboratory, RIT, 1-69.

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

_____ and W.G. Blandford (1924), *A Grammar of Spoken English on a Strictly Phonetic Basis,* Cambridge: W. Heffer and Sons.

Piaget, J. (1959), *The Language and Thought of the Child,* London: Routledge and Kegan Paul.

Pike, E.G. (1949), "Controlled Infant Intonation," *Language Learning, 2,* 21-24.

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

_____(1965), "On the Grammar of Intonation," *Proc. 5th Int'l. Cong. Phon. Sci.*, Basel, 105-117.

Pittenger, R.E., Hockett, C.F., and J.J. Danehy (1960), *The First Five Minutes*, Ithaca, N.Y.

Pollack, I. (1952), "The Information of Elementary Auditory Displays," *J. Acoust. Soc. Am.*, *24*, 745-749.

_____(1953), "The Information of Elementary Auditory Displays II," *J. Acoust. Soc. Am.*, *25*, 765-769.

_____ and L. Ficks (1954), "Information of Elementary Multi-Dimensional Auditory Displays," *J. Acoust. Soc. Am.*, *26*, 155-158.

_____ and J.M. Pickett (1964), "The Intelligibility of Excerpts from Conversation," *Language and Speech*, *6*, 165-171.

Postal, P. (1968), *Aspects of Phonological Theory*, New York: Harper and Row.

Quirk, R., Duckworth, A.P., Svartik, J., Rusiecki, J.P.L., and A.J.T. Colin (1964), "Studies in the Correspondence of Prosodic to Grammatical Features in English," *Proc. 9th Int'l. Cong. Ling.*, The Hague: Mouton, 679-691.

Rigault, A. (1962), "Rôle de la frequence, de l'intensité et de la durée vocaliques dans la perception de l'accent en français," *Proc. 4th Int'l Cong. Phon. Sci.*, The Hague: Mouton, 735-748.

Ringel, R.L. and D.D. Kluppel (1964), "Neonatal Crying: A Normative Study," *Folia Phoniatrica*, *16*, 1-9.

Risberg, A. (1962), "Fundamental Frequency Tracking," *Proc. 4th Int'l. Cong. Phon. Sce.*, The Hague: Mouton.

Schubiger, M. (1935), *The Role of Intonation in Spoken English*, St. Gallen.

_____(1958), *English Intonation: Its Form and Function*, Tübingen, Max Niemeyer Verlag.

Sharp, A.E. (1958), "Falling-rising Intonation Patterns in English," *Phonetica*, *2*, 127-152.

Sheppard, W.C. and H.L. Lane (1968), "Development of the Prosodic Features of Infant Vocalizing," *J. Speech and Hear. Res.*, *11*, 94-108.

Sledd, J. (1955), Review of G.L. Trager and H.L. Smith, Jr. (1951), *Outline of English Structure*, *Language*, *31*, 312-335.

Smith, H.L., Jr. (1955), Review of W. Jassem (1952), *Intonation of Conversational English*, *Language*, *31*, 150-153.

Stevick, R.D. (1963), "The Biological Model and Historical Linguistics," *Language*, *39*, 159-169.

Stockwell, R.P. (1960), "The Role of Intonation in a Generative Grammar of English," *Language 36*, 360-367.

_____(1962), "On the Analysis of English Intonation," Second Texas Conference on Problems of Linguistic Analysis in English: *Studies in American English*, Austin, Texas, 39-55.

_____(1963), Review of D.L. Bolinger (1961a), *Generality, Gradience, and the All-or-None*, *Language*, *39*, 87-91.

Studdert-Kennedy, M. and K. Hadding-Koch (1969, forthcoming), "Further Experimental Studies of Fundamental Frequency Contours."

Sweet, Henry (1892), *New English Grammar*, Pt. 1, Oxford: Clarendon Press.

Templin, M.C. (1957), *Certain Language Skills in Children: Their Development and Interrelationships*, Minneapolis: Univ. of Minn. Press.

Torgerson, W.S. (1958), *Theory and Methods of Scaling*, New York, Wiley.

Trager, G.L. and H.L. Smith, Jr. (1951), *Outline of English Structure* (Studies in Linguistics, Occasional Papers, 3), Norman, Oklahoma.

Uldall, E. (1960), "Attitudinal Meanings Conveyed by Intonation Contours," *Language and Speech*, *3*, 223-234.

_____(1962), "Ambiguity: Question or Statement? or 'Are You Asking Me or Telling Me?'," *Proc. 4th Int'l. Cong. Phon. Sci.*, The Hague: Mouton.

_____(1964), "Dimensions of Meaning in Intonation," in D. Abercrombie *et. al.*, eds., *In Honour of Daniel Jones*, London: Longmans, 101-112.

Vanderslice, R. (1967), "Larynx vs. Lungs: Cricothyrometer Data Refuting some Recent Claims concerning Intonation and 'archetypality'," *Working Papers in Phonetics*, *7* (UCLA Phonetics Laboratory) 69-79.

_____(1968), *Synthetic Elocution*, *Working Papers in Phonetics*, *8* (UCLA Phonetics Laboratory).

_____(1969), "Intonation, Scientism, and 'Archetypality'," (Review of P. Lieberman, *Intonation, Perception, and Language*), in *Studies in Language and Language Behavior*, Progress Report 8, Center for Research on Language and Language Behavior, Univ. of Mich.

Velten, H.V. (1943), "The Growth of Phonemic and Lexical Patterns in Infant Language," *Language, 19,* 281-292.

Wang, W. S-Y. (1962), "Stress in English," *Language Learning, 12,* 69-77.

_____(1967),"Phonological Features of Tone," *IJAL, 33,* 93-105.

Weir, R.H. (1962), *Language in the Crib,* The Hague: Mouton.

_____(1966), "Some Questions on the Child's Learning of Phonology," in *The Genesis of Language* (eds. F. Smith and G.A. Miller), Cambridge: MIT Press, 153-168.

Wells, R.S. (1947), "Immediate Constituents," *Language, 23,* 81-117.

Winer, B.J. (1962), *Statistical Principles in Experimental Design,* New York: McGraw-Hill.

Zwirner, E. (1932), "A Contribution to the Theory of Pitch Curves," *Archives Néerlandaises de Phonétique Expérimentale, 7,* 38-51.

Zwirner, E. (1952), "Probleme der Sprachmelodie," *Zeitschrift für Phonetik, 6,* 1-12.