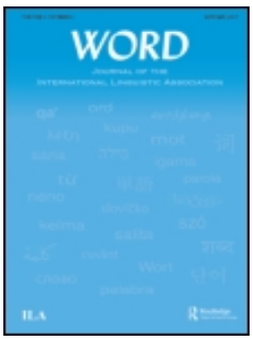

Word

ISSN: 0043-7956 (Print) 2373-5112 (Online) Journal homepage: https://www.tandfonline.com/loi/rwrd20

On the Acoustic Basis of the Perception of Intonation by Linguists

Philip Lieberman

To cite this article: Philip Lieberman (1965) On the Acoustic Basis of the Perception of Intonation by Linguists, Word, 21:1, 40-54, DOI: 10.1080/00437956.1965.11435417

To link to this article: https://doi.org/10.1080/00437956.1965.11435417

Published online: 16 Jun 2015.

PHILIP LIEBERMAN

On the Acoustic Basis of the Perception of Intonation by Linguists

Several different systems for the transcription of intonation are in wide use today. We will use the term intonation throughout this paper to mean the entire ensemble of pitch contours, pitch levels, and stress levels that occurs when a sentence is spoken. "Tonetic" transcriptions such as those described by Sweet,[1] Kindon,[2] Jassem,[3] and Stockwell,[4] are used by many linguists. These systems do not imply a sharp dichotomy between stress and pitch. Essentially they treat stress and pitch as related elements and note whether the stressed syllables of an utterance have level, rising or falling pitch contours. Other linguists,[5] who also believe that stress and pitch are related, prefer to transcribe the total intonation contour in terms of "tunes." Jones,[6] for example, employs two basic tunes to describe the "essential" elements of English intonation. The tunes are atomistic ensembles of stressed and unstressed syllables and pitch contours.

Still other linguists, however, prefer to use systems that yield transcriptions in which pitch and stress are completely independent "phonemic" entities. One such system is that described by Trager and Smith in 1951 in their *Outline of Linguistic Analysis*. This system is in wide use. It also makes perhaps the most detailed claims regarding the function of pitch and stress in language. We shall describe in this paper an experiment that was designed to ascertain what aspects of the acoustic signal linguists actually note when they make Trager–Smith transcriptions. We will also discuss some aspects of other systems for the transcription of intonation insofar as they relate to the data of this experiment.

[1] *New English Grammar* Part I (Oxford, 1892), p. 228.

[2] "Tonetic Stress Markers for English," *Le Maître Phonetique* III, LIV (1939), 60–64.

[3] *Intonation of Conversational English* (Wrocław, 1952).

[4] Review of Kindon's *The Groundwork of English Intonation* (1958) in *International Journal of American Linguistics* XXVII (1961), 278ff.

[5] L. E. Armstrong and I. C. Ward, *Handbook of English Intonation* (Leipzig and Berlin, 1926).

[6] D. Jones, *An Outline of English Phonetics*, 3rd Ed. (New York, 1932).

The Trager–Smith system makes use of four pitch levels, three terminal junctures, and various vocal qualifiers to describe the pitch contour of an utterance. It also uses four levels of stress to describe the stress relationships of the speech signal. Stress and pitch are supposed to be completely independent. The linguist is, for example, supposed to be able to perceive the stress levels of an utterance independently of his perception of the pitch levels.

Stress and pitch are supposed to relate to rather distinct linguistic levels in this system. Stresses are supposed to distinguish certain morphemic classes from each other while pitch levels and terminals are supposed to provide acoustic cues that tell a listener where the phrases of a sentence begin or end. In the words of the *Outline of English Structure* (page 77):

> The contribution of the phonological analysis of stress, juncture, and intonation patterns . . . is that it makes . . . the recognition of immediate constituents and part of speech syntax into solidly established objective procedures, removing once and for all the necessity of defending one's subjective judgements as to what goes with what.

The object of this experiment was to test whether the linguist using Trager–Smith notation does in fact employ an "objective" procedure in which he considers the physically present acoustic signal. The results of the experiment will demonstrate that the linguist often considers his "subjective" judgement and fills in the Trager–Smith pitch notation that is appropriate to the structure of the sentence, which he usually infers from the words of the sentence and his knowledge of the language.

PROCEDURE

The procedure of this experiment involved the use of a set of emotional and non-emotional utterances that have been used for experiments[7] designed to see how much information the modulation of the fundamental frequency and amplitude of a speaker's voice conveys. Naive, phonetically untrained listeners were used for the previous experiments.

In those experiments a group of male native speakers of American English read a set of eight neutral sentences in an anechoic chamber. Each speaker was instructed to read the sentences with appropriate vocal modifications so that they could be identified as belonging to one of the following eight categories or emotional modes: (1) a bored statement, (2) a confidential communication, (3) a message expressing disbelief or doubt, (4) a

[7] P. Lieberman, "Perturbations in Vocal Pitch," *Journal of the Acoustical Society of America* XXXIII (1961), 597–703; P. Lieberman and S. D. Michaels, "Some Aspects of Fundamental Frequency, Envelope Amplitude and the Emotional Content of Speech," *Journal of the Acoustical Society of America* XXXIV (1962), 922–927.

message expressing fear, (5) a message expressing happiness, (6) an objective question, (7) an objective statement, and (8) a pompous statement. Each sentence was read three times in each mode.

The 24 repetitions of each sentence were then placed in a random sequence and categorized by a group of 20 untrained listeners in a forced judgement test to select the most identifiable utterance from each of the eight categories for each sentence. A panel of trained observers then listened to the same set and rejected those utterances that they found to be strained or unnatural. On the basis of these criteria the best utterances of each of the eight emotional categories for each of the eight sentences of each speaker were selected.

For the present experiment the utterances of two of the speakers reading the sentence, *They have bought a new car*, were selected. The pitch of the utterances was electronically extracted as was the amplitude of the speech signal's envelope. These control signals were used to generate stimuli on a fixed POVO[8] which produced [a]'s whose fundamental frequency and envelope amplitude varied in the same manner as the original speech signals. We also smoothed the fundamental frequency contours and produced [a]'s on the POVO which lacked rapid pitch perturbations or variations, as well as [a]'s in which we did not modulate the amplitude of the POVO output. In all, stimuli were produced in which:

a. Fundamental frequency and envelope amplitude information were preserved,
b. only fundamental frequency information was preserved,
c. only smoothed fundamental frequency information was preserved,
d. only amplitude information was preserved.

Preparation of Test Stimuli

Electronic circuits were used to derive a marker pulse on the leading edge of the amplitude peak of each fundamental period of the recorded speech samples. These measurements of the fundamental frequency contours of the speech signal were accurate to within 0.2 milliseconds of the duration of each fundamental period. These pulses could be then used to drive the POVO fixed vowel synthesizer which had formant frequencies of 750, 1100, and 2450 cycles per second and bandwidths of 70, 80, and 115 cps respec-

[8] A POVO is an electronic analog of the vocal tract that employs a series of resonant circuits which correspond to the formant frequencies or natural resonances of the pharyngeal cavity, mouth, tongue, lips, etc. The fixed POVO when it was excited by a series of pulses which approximate the acoustic output of the vocal cords produced a vowel-like sound, cf. K. N. Stevens "Synthesis of Speech by Electronic Analog Devices," *Journal of the Audio Engineering Society* IV (1956), 2–8.

tively. The output waveform of the POVO was also made asymmetrical to furnish a better approximation to human speech.

A tape recording was made of the POVO's output when it was excited by these pulses. The tape recording consisted of a number of short utterances. Each utterance consisted of a series of [a]'s that had the same fundamental frequency contour as one of the sentences that the speakers read. The temporal pattern of these [a]'s was identical to that of the voiced portions of the sentence that the speaker read. The [a]'s on this first tape recording had constant amplitudes and the words of the sentence, of course, were not present. This essentially isolated all of the fundamental frequency information of the original speech signal, removing all phonetic and amplitude information.

A second tape recording was then made in which a 20 msec full wave rectifying circuit was used to obtain the envelope amplitude of the original speech signal and modulate the POVO so that the synthesized waveform retained part of the amplitude information present in the original speech signal. The utterances on this second tape recording therefore consisted of sequences of [a]'s that had the same fundamental frequencies, envelope amplitudes and temporal pattern as the sentences that the speakers read.

A third tape recording was made in which the amplitude modulation of the POVO was retained but where the short term variations in the fundamental frequency contours of the [a]'s were electronically smoothed out. The smoothing time constant was 40 msec. The fundamental frequency contours of the [a]'s on this tape recording had the same range and shape as the fundamental frequency contours of the original speech signal. Many of the short, rapid perturbations or variations in fundamental frequency that occur during normal speech were however smoothed out of the stimuli on this tape recording. A fourth tape was also made in which the smoothing time constant was increased to 100 msec.

A fifth tape recording was made in which the POVO was excited by a constant 120 cps pulse source so that the output waveform had the same temporal pattern and envelope amplitude as the voiced portions of the original speech signal. This tape recording isolated amplitude information from the fundamental frequency and phonetic information of the sentences that the speakers read.

Measurement of Test Stimuli

The contours of the fundamental frequency as a function of time were automatically plotted for all of these stimuli by means of a computer program which displayed the contours on the oscillographic output of a PDP-1 digital computer. The oscilloscope was photographed and the

photographs were then enlarged. The fundamental frequency could then be measured from these plots with a resolution of approximately 5 cps.

The Listening Tests

Two linguists[9] who were quite familiar with the Trager–Smith system transcribed these stimuli. We would mail each linguist two tape recordings in which all sixteen sentences were presented under two processing conditions. The linguists transcribed the tape recordings using the Trager–Smith system. One linguist also made free hand drawings of the pitch contours. This same linguist also transcribed the utterances using a "tonetic" system. Each linguist independently transcribed the set of tape recordings and mailed the recordings and the transcriptions back. We then would mail back two more tape recordings after a few months had elapsed. At the end of the experiment, which took about two years, the linguists transcribed the complete sentences without any electronic processing so that they heard the words of the sentences as well as the fundamental frequency and amplitude information.

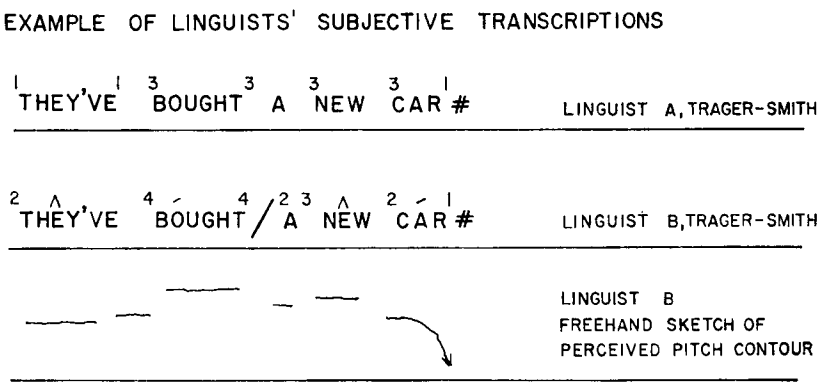

FIGURE 1

[9] The author would like to acknowledge the aid of Dr. Leigh Lisker of the University of Pennsylvania and Dr. Robert P. Stockwell of the University of California, Los Angeles, who transcribed the speech material that forms the basis of this experiment. Part of this material was presented at the Sixty-Seventh Meeting of the Acoustical Society of America, 9 May 1964, New York City.

Observations and Discussion

In Figure 1 we have an example of the two linguists' transcriptions of the same sentence when they heard all the words of the sentence. The superscribed numbers refer to the Trager–Smith phonemic pitch levels. Pitch level 1 is defined as the lowest pitch while pitch level 4 is the highest pitch. Several terminal junctures are also used. The symbol /, single bar, refers to a sustention of pitch. The symbol #, double cross, refers to a falling pitch. A third juncture //, double bar, refers to a rising pitch. Linguists A's and B's transcriptions did not agree. Linguist A used pitch level 1 at the start of the sentence where linguist B used pitch level 2. Linguist A used pitch level 3 at the beginning of *bought* where linguist B used pitch level 4. The extra instance of pitch level 1 at the end of *they've* by Linguist A merely represents the continuation of the pitch level at the start of *they've* and really represents the initial pitch level that has been continued. The phonemic pitch levels and terminal symbols that the two linguists used when they transcribed unprocessed sentences where they clearly heard the words differed 60 percent of the time, even after nonsignificant reiterations of pitch levels and terminal symbols were discounted.

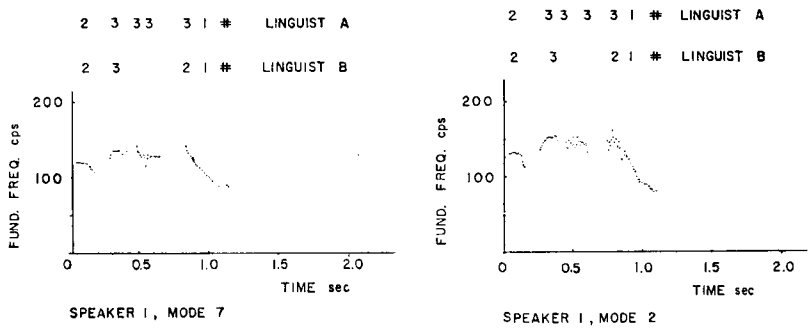

Figure 2

Only certain limited classes of utterances were transcribed with reasonable consistency by the two linguists. In Figure 2, two of these utterances' pitch transcriptions are presented together with the actual contour of fundamental frequency plotted with respect to time. If this study had been limited to contours like those plotted in Figure 2 we might have concluded that the linguists' transcriptions were reasonably accurate representations of the actual frequency contour of the speech signal. Note that both of the contours in Figure 2 rise slightly from a starting frequency and then fall at

the end of the sentence. This contour is quite commonly encountered in American English. Pike,[10] for example, notes that falling contours constitute 62.5 per cent of the contours in a "short exposition" and that 72 per cent of the sentences conclude with falling contours. Trager and Smith[11] note that the commonest pitch morpheme in American English is 231#. Daniel Jones[12] also notes that his contour is quite common in American English. However, he considers this contour as an example of Tune I.

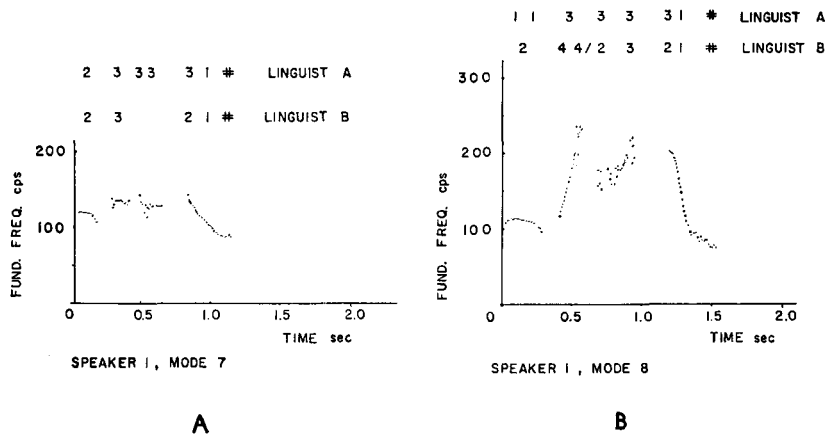

FIGURE 3

In Figure 3A we again see one of the fundamental frequency contours that appeared in Figure 2 in addition to another contour (Figure 3B) which involves rather different fundamental frequencies. Note that the second part of the contour (in Figure 3B) bears the same Trager–Smith transcription as does the complete contour in Figure 3A. This segment of the contour in Figure 3B which is transcribed as 2321# by Linguist B is an example of Tune I, as is the entire contour in Figure 3A. Both contours rise from a starting point and then fall at the sentence's end.

The linguists apparently "hear" both of these signals as identical stimuli though rather different fundamental frequencies are actually involved. The stimulus is recognized as an entity and then recorded in terms of the pitch levels appropriate to the most common "suprasegmental morpheme" of

[10] *The Intonation of American English* (Ann Arbor, 1945), pp. 155–157.
[11] *Op. cit.*
[12] *An Outline of English Phonetics* 9th Ed. (Cambridge, 1962), p. 362.

American English. This effect was quite common. The linguists often transcribed contours that involved rather different fundamental frequencies with similar pitch levels.

If the linguists really were transcribing the intonation in terms of free "phonemic" pitch levels it would be reasonable to find some acoustic correlates of the pitch levels. We might, for example, find ranges of fundamental frequencies that corresponded to each pitch level. Since the average fundamental frequencies of different speakers varied, these ranges probably would be relative to the average fundamental frequency of a given speaker. Pitch level 1 for a woman, for example, might correspond to a higher fundamental frequency range than pitch level 1 for a man's voice. Hopefully, all linguists would transcribe the same ranges of fundamental frequencies as the same pitch levels for a particular speaker's utterances. However, there is so much uncertainty about the perception of intonation[13] that we might well find that different linguists had different relative ranges.

We therefore measured the fundamental frequency that corresponded to each pitch level entry of the linguists' transcriptions of the clear, unprocessed speech samples. This posed no special problems except for the pitch levels that occurred in utterance-final position before the terminal juncture. In these cases the contour often ended with a rather rapid fundamental frequency transition which must be correlated with the terminal juncture. The fundamental frequency of the last pitch level was therefore measured at a point that was 30 per cent of the difference between the fundamental frequency of the previous pitch level of the final syllable and the fundamental frequency with which the contour ended.

In Figure 4 we have plotted the fundamental frequencies that corresponded to each pitch level transcribed by the linguists except for those pitch levels that were part of sentences marked with the vocal qualifiers (overhigh, overlow, etc.) by either linguist. The pitch levels that were measured were therefore not influenced by the presence of vocal qualifiers. The ordinate bears the scale of fundamental frequencies in cps. The *o's* represent the entries for the transcriptions of speaker one's utterances and the *x*'s the entries for speaker two. Note that although there is a distinct upward trend for the fundamental frequencies associated with increasing pitch levels, it is rather difficult to regard the pitch levels as representing discrete absolute levels even when we restrict the comparison to one

[13] J. Sledd, in a review of the *Outline of English Structure* in *Language* XXXI (1955), page 316, notes that, "Anyone who has attempted to analyze or teach the English patterns of pitch and stress knows that competent observers may vigorously disagree and a single observer may disagree with himself so often as to make secure confidence in his own judgments painfully difficult. . . ."

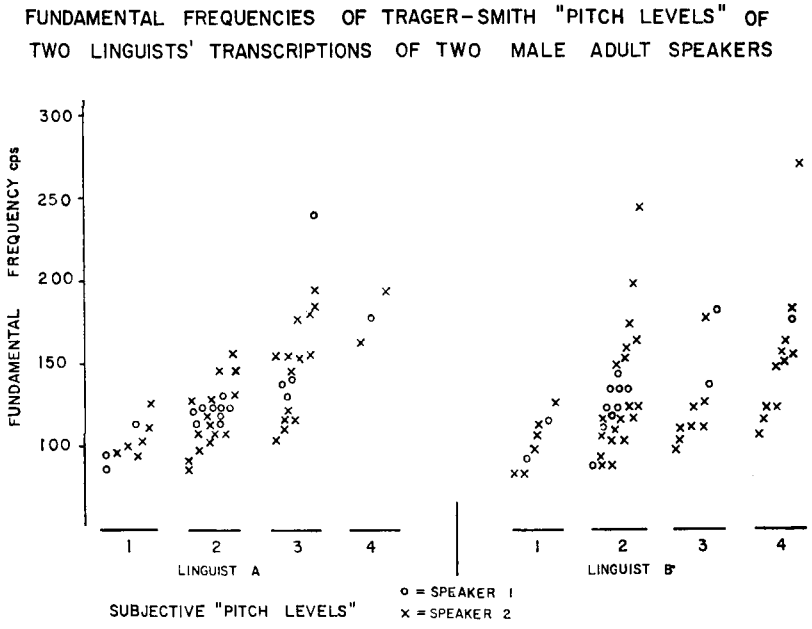

FIGURE 4

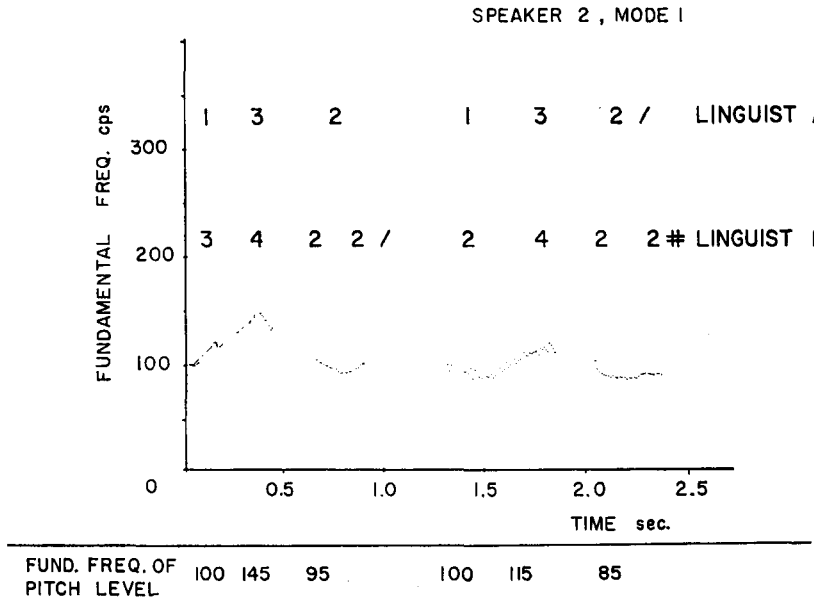

FIGURE 5

linguist's transcriptions of one speaker's utterances. We might still, however, ask whether the pitch levels represented the linguist's perception of relative fundamental frequencies.

One of the least general of all the possible conditions under which the pitch levels might represent relative fundamental frequencies simply requires that the fundamental frequencies corresponding to the pitch levels be monotonically increasing for the transcriptions of a single speaker's utterances by a single linguist.

In Figure 5 we have the actual contour of fundamental frequency plotted with respect to time for a single utterance. We also have the two transcriptions by the two linguists. Note that even within this single utterance pitch level 1 may be approximately 100 cps while pitch level 2 may range from 85 to 95 cps for linguist A. Similar results were noted for linguist B's transcriptions. The fundamental frequencies associated with lower pitch levels may or may not be lower than those associated with higher pitch levels. The pitch levels do not appear to represent relative fundamental frequencies with any reasonable degree of certainty. The most we can say is that within a given segment of continuous voicing the fundamental frequencies that correspond to the pitch levels are monotonically increasing as we go from low to high pitch levels.

Comparison of Transcriptions of POVO Stimuli with Complete Speech Signal

Let us now consider the transcriptions that the linguists made of the POVO stimuli in which the words of the sentences were effectively removed from the speech signal, leaving only fundamental frequency and amplitude contours. The POVO produced a fixed vowel-like [a] whose amplitude and fundamental frequency varied. We first compared the transcriptions of the complete sentences with those of the POVO stimuli that had the same frequency and amplitude contours as the original speech signal (tape recording two). We found that each linguist changed fifty percent of the pitch levels and junctures of his transcription when he heard the POVO stimuli on tape recording two. Furthermore when the transcriptions of these POVO stimuli were compared with the actual contours of fundamental frequency as a function of time we found that the linguists' transcriptions of the POVO stimuli were more accurate than their transcriptions of the unprocessed stimuli where they heard the sentences' words.

In Figure 6 the transcriptions of linguist B for the complete sentence and the POVO stimulus are presented along with the actual fundamental frequency contour. Note that the transcription of the stimulus generated with the POVO is considerably more accurate than is the transcription made in the clear. The linguists' ears were remarkably good so long as they

did not hear the words of the message. They accurately tracked the fundamental frequency contours when they listened to the synthesized stimuli in test tapes three and four where the contours were smoothed. The linguists also correctly transcribed the monotones recorded on tape recording five.

The pitch transcriptions of the POVO stimuli with fundamental frequency and amplitude information and the POVO stimuli with only fundamental frequency information differed only 15 percent of the time. This of course, indicates that fundamental frequency is closely related to the linguist's perception of pitch.

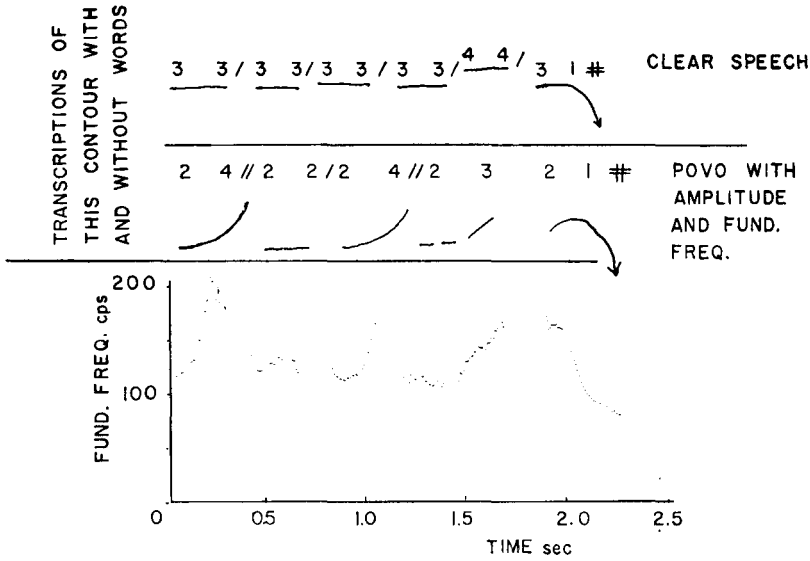

FIGURE 6

The Perception of Phonemic Stress Levels

Linguist B transcribed four degrees of stress in addition to the four pitch levels and three junctures of the Trager–Smith system. Referring back to Figure 1 the symbol ˈ represents a primary stress, while the symbol ˆ, the circumflex, represents a secondary stress. The linguist transcribed 34 primary, 27 secondary and 3 tertiary stresses for this set of utterances. When he transcribed the POVO stimuli synthesized with both fundamental frequency and amplitude information (tape recording two) 66 percent of his primary stresses were unchanged but only 7 percent of his secondary

stresses were preserved. Several independent experiments[14] have demonstrated that the acoustic correlates of stress include fundamental frequency, amplitude and duration. All of these acoustic correlates were preserved in the POVO stimuli. The loss of the secondary and tertiary stresses suggests that the linguist may be inferring the presence of these stresses from his knowledge of the grammatical attributes of the words of the sentence rather than hearing them in the physical attributes of the speech signal.

The stress assignments in linguist B's transcription of tape recording one (where the amplitude information of the original speech signal was not present) differed from his transcription of tape recording two only 10 percent of the time. This is also consistent with the findings of these independent experiments which have, in general, noted that fundamental frequency is a stronger cue for the perception of stress than the amplitude of the speech signal.

Tonetic Transcriptions of Intonation

One of the linguists also transcribed the entire stimulus ensemble using a tonetic notion.[15] This transcription noted whether the pitch of stressed syllables was high or low and whether it was level, rising or falling (or rising–falling, falling–rising, etc.). The transcription also noted the presence of two junctures, | and ↓, where the former occurred with any dynamic accent and the latter only with falling ones. Extra emphasis also could be indicated by this system. This notation makes far fewer claims about the details of the linguist's perception of intonation and stress than does the Trager–Smith system. It implicitly states that the perception of pitch at any given instant is always relative to the entire intonation contour. It also states that the perception of pitch and stress is closely related.

We found that this notation was more consistent than the Trager–Smith transcriptions when we compared the transcriptions of the complete sentences with the transcriptions of the POVO stimuli. The linguist changed only 25 percent of his notation when he heard tape recording two, in which the words of the sentences were removed, leaving only fundamental frequency and amplitude information. Most of the errors involved extra occurrences of the juncture |. However, fewer errors occurred with respect to this juncture than with respect to Trager–Smith / and //. It seems likely

[14] D. B. Fry, "Duration and Intensity as Physical Correlates of Linguistic Stress," *Journal of the Acoustical Society of America* XXXII (1955), 765–769. D. B. Fry, "Experiments in the Perception of Stress," *Language and Speech* (1958), 126–152. P. Lieberman, "Some Acoustic Correlates of Word Stress in American English," *Journal of the Acoustical Society of America* XXXII (1960), 451–454.

[15] R. Stockwell, *op. cit.*

that a single "tentative" juncture exists whose acoustic correlate is a "not-falling" fundamental frequency contour. Pike,[16] of course, uses only one "tentative" pause while Jones[17] and Armstrong and Ward[18] use Tune II for this purpose. Hadding-Koch[19] in a study of the intonation of Southern Swedish also found that listeners were able to perceive only two types of junctures, "non-final" and "final."

Summary of Observations and Discussion

1. We found that when two competent linguists independently transcribe a set of sentences that include "emotional" as well as "normal" utterances 60 percent of the pitch levels and junctures of the two Trager–Smith transcriptions vary.

2. The Trager–Smith pitch levels do not correspond to discrete non-overlapping ranges of fundamental frequency nor do they correspond to discrete relative ranges of fundamental frequency. These comments apply even when we consider the transcriptions made by a single linguist who carefully transcribed the tape recorded sentences of a single talker.

3. The pitch levels of the Trager–Smith system do not even reflect the relative pitch levels of a single utterance of a single talker when it is transcribed by a single linguist. The fundamental frequency that corresponds to pitch level one, for example, may be identical to or greater than the fundamental frequency that corresponds to pitch level two. The pitch levels reflect the relative fundamental frequency only during segments of speech in which there is continuous voicing. The "phonemic" pitch levels at best therefore do not provide any more information than does a "tonetic" transcription.

4. A subclass of utterances were transcribed more consistently and accurately than the rest of the stimulus ensemble. These utterances were characterized by what Jones has termed a single instance of Tune I or Tune II. These contours were always transcribed by the linguists in terms of the "supra-segmental morphemes" $\sqrt{231\#}$, $\sqrt{232/}$, or $\sqrt{232//}$, etc. When the single contour extended over an entire unemotional utterance the pitch levels bore a reasonable relationship to the actual fundamental frequency contour of the utterance. However, in other instances, contours having the same "shape" but different fundamental frequency ranges were transcribed with exactly the same pitch levels and junctures. The linguists

[16] Pike, *op. cit.*, 31–32.

[17] Jones, *An Outline of English Phonetics*, 3rd Ed. (New York, 1932), pp. 350ff.

[18] Armstong and Ward, *op. cit.*

[19] K. Hadding-Kock, *Acoustico-Phonetic Studies in the Intonation of Southern Swedish* (Lund, 1961), p. 61.

apparently responded to the general form of the contour rather than to any pitch levels. These effects were quite general and occurred when these contours encompassed only part of an utterance. We are excluding from this discussion all of the utterances marked with any vocal qualifiers, e.g. "overhigh."

5. The linguists heard stimuli in which the fundamental frequency and amplitude contours of the complete sentences were accurately reproduced as modulations of a fixed vowel. When the linguists transcribed these contours we found that each linguist changed fifty percent of the pitch levels and junctures of his transcription vis-a-vis his transcription of the complete sentence where he, of course, heard the words of the message.

The transcriptions of the fixed vowel were more accurate representations of the actual fundamental frequency contours than the transcriptions of the complete speech signal where the linguists heard the words of the message.

6. When the linguist heard the complete speech signal he was able to transcribe four degrees of stress. However, when the linguist heard the fixed vowels that were accurately modulated with the fundamental frequency and amplitude contours of the original speech signal, he was unable to transcribe accurately more than two degrees of stress, stressed or unstressed. Only 7 percent of the secondary stresses and none of the tertiary stresses that were transcribed for the complete speech signal were transcribed under these conditions. These results suggest that only two degrees of stress may have acoustic correlates independent of vowel quality. A vowel may be either stressed or unstressed (vowel reduction phenomena may add a third degree of stress).

Conclusion

In conclusion, the results of this experiment suggest that the phonemic pitch levels and terminal symbols of the Trager–Smith system often have no distinct physical basis. The linguist infers their presence from his knowledge of the transcriptions that the Trager–Smith system usually uses for certain combinations of words. The same comments seem to apply to secondary and tertiary stresses. Moreover, the results of this experiment indicate that there is no basis for regarding the Trager–Smith pitch levels as the perceptual manifestations of either absolute or relative fundamental frequency ranges except for certain contours that recur quite frequently in normal discourse. However, these contours appear to be perceived as complete entities. When other intonation contours occur, the Trager–Smith notation becomes inconsistent and has no reasonable relationship to those attributes of the physical signal which it supposedly is transcribing.

An independently motivated generative model[20] shows that the intonation of a sentence can be predicted if one considers three sets of factors: (1) the physiological constraints imposed by the human respiratory system, (2) the emotional state of the speaker, and (3) the ultimate recoverability of the Deep Phrase Marker[21] that underlies the final phonological shape of the sentence. The generation of the intonation of an utterance is organized, in part, in terms of certain synchronized patterns of the muscular activity of the larynx and the respiratory system. Perception, in this model, involves "analysis by synthesis".[22] Intonation is therefore perceived in terms of complete contours of fundamental frequency and amplitude, i.e., ensembles of fundamental frequency functions and amplitude variations as functions of time. The data of this experiment support this generative model.

Air Force Cambridge Research Laboratories
Bedford, Massachusetts 0731

[20] P. Lieberman, "Intonation and the Syntactic Processing of Speech", in *Proceedings of the Symposium on Models for Perception of Speech and Visual Form*, 11–14 November 1964, Boston, Massachusetts.

[21] P. Postal and J. Katz, *An Integrated Theory of Linguistic Description* (1964).

[22] M. Halle and K. N. Stevens, "Analysis by Synthesis," in *Proceedings of the Seminar on Speech Compression and Processing*, 28–30 September 1959, Bedford, Massachusetts. AFCRC-TR-59–198.