Paralinguistic Use of Speech Melody

Yukio Takefuta, Ph.D.
Chiba University, Chiba, Japan
and
Elizabeth G. Jancosek, Ph.D.
University of Colorado, Boulder, U.S.A.

Introduction

When we communicate with other people by using a language, we use speech sounds. The sounds are used to represent the symbols of the language and transmit them to the listeners. No one, with the possible exception of linguists, really tries to count the number of different sounds which are used in his mother tongue. When we hear a foreign language it sounds as if every sound that reaches our ears is a new one. However, the total number of segmental sounds which most languages use is actually limited to 40 or 50 if only linguistically significant sounds, i.e. the sounds by which we can distinguish meanings (phonemes), are counted separately. For example, English has 44 phonemes (Dewey, 1923) while Japanese has only 23 (Takebayashi, 1970).

It is well known that the segmental phonemes can be classified into vowels and consonants. It is less obvious to the average speaker that they can also be classified into other categories: voiced and unvoiced. The latter classification is more significant for the discussion of the present topic. According to a study by Dewey, 78 per cent of the sounds used in standard English prose are voiced, / ɪ, ə , æ , ɛ , ɒ , ʌ , i, e, eɪ , u, aɪ , ou, ɔ , ʊ, aʊ, ɑ, o, ju, ɔɪ , n, r, d, l, ð , z, m, v, w, b, ŋ , g, j, dʒ , ʒ / and 22 per cent are unvoiced / t, s, k, p, f, h, ʃ , tʃ , θ /. Voelker, Denes and Roberts found similar ratios in studies conducted more recently (Voelker, 1935; Denes, 1963; Roberts, 1965).

In normal situations, i.e., excluding whispered speech, voiced sounds are produced with the vocal folds vibrating and unvoiced ones, without vibration. Thus, when voiced sounds are produced, the concomitant sound waves show periodic waveforms and we tend to hear musical tones from them. Since voiced sounds are distributed more or less evenly through the entire span of any utterance there is an accompanying melody throughout as demonstrated in the sentence below;

Baseball is very popular in America.
/ beɪsbọl ɪz vẹrɪ pọpjụlẹr ɪn ẹmẹrɪkẹ /
(Sounds with dots below are voiced.)

In our communicative behavior, speakers use this melody intentionally or unintentionally, to convey certain messages, most of which no words or grammatical structures can express.

In this paper, we will discuss the types of messages which can be conveyed by such melodies.

Speech Melody

All voiced sounds, especially vowels, have characteristic frequencies with which the sounds are produced if no special effort is made to change the frequency (Black, 1949; and others). We perceive a higher pitch from a higher rate of vocal fold vibration (or frequency) and a lower pitch from a vibration of a lower rate. Thus, theoretically, the melodies or the pitch patterns,

of the utterances may differ depending upon the kinds and the order of the constituent voiced sounds. The pitch pattern, however, varies more significantly for several other reasons. First, because of the normal physical constraints of a vibrating body, the vocal folds start vibrating at a lower rate, gradually increase in rate and finally, due to inertia the rate decreases gradually before vibration ceases. Thus, if completely colorless, the natural frequency contour of an utterance drawn as a function of time would appear as the curve in Figure 1.

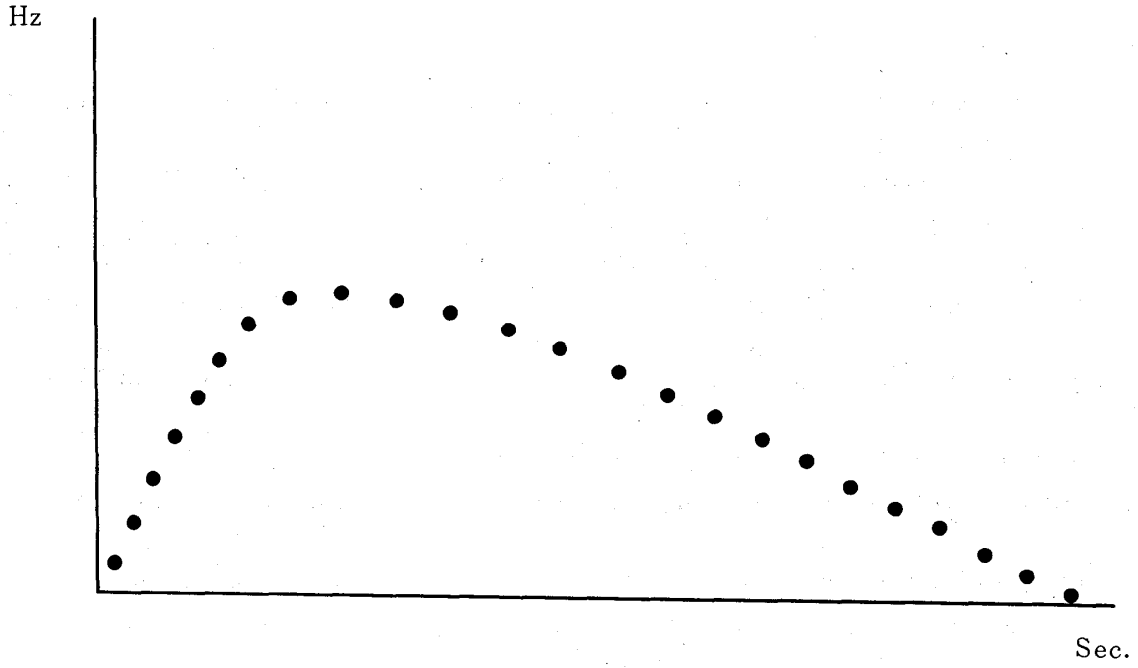

Figure 1. A frequency contour to be expected if no special effort is made to alter the natural physiological constraints of the activity of the vocal chords.

This curve is exclusive of the effects of the characteristic frequencies of voiced sounds; it represents the basic physiological function of phonation. Such a pattern of frequency changes plotted as a function of time is called a "melody curve" or a "frequency contour". The accompanying auditory impression may be termed a "speech melody".

Almost no pattern of an actual utterance will look as smooth and simple as the one in Figure 1. We requested a group of American college students to pronounce a sentence with as many different intonation patterns as they could imagine. They produced six different pitch patterns on the average. The number is the average of the attempts made by 60 speakers for each of 10 everyday sentences. All of the patterns of the "frequency contour" obtained from one speaker for the sentence "I don't know" are shown in Figure 2. The speaker wrote the intended message associated with each of the intonation patterns produced. The intended messages were: "unsure", "answer", "defense", "anger", and "pretence".

Speakers in the above survey were producing the intonation patterns in an artificial situation in a sound-treated booth. It is quite likely that the speakers could have produced more varieties of intonation patterns in real situations.

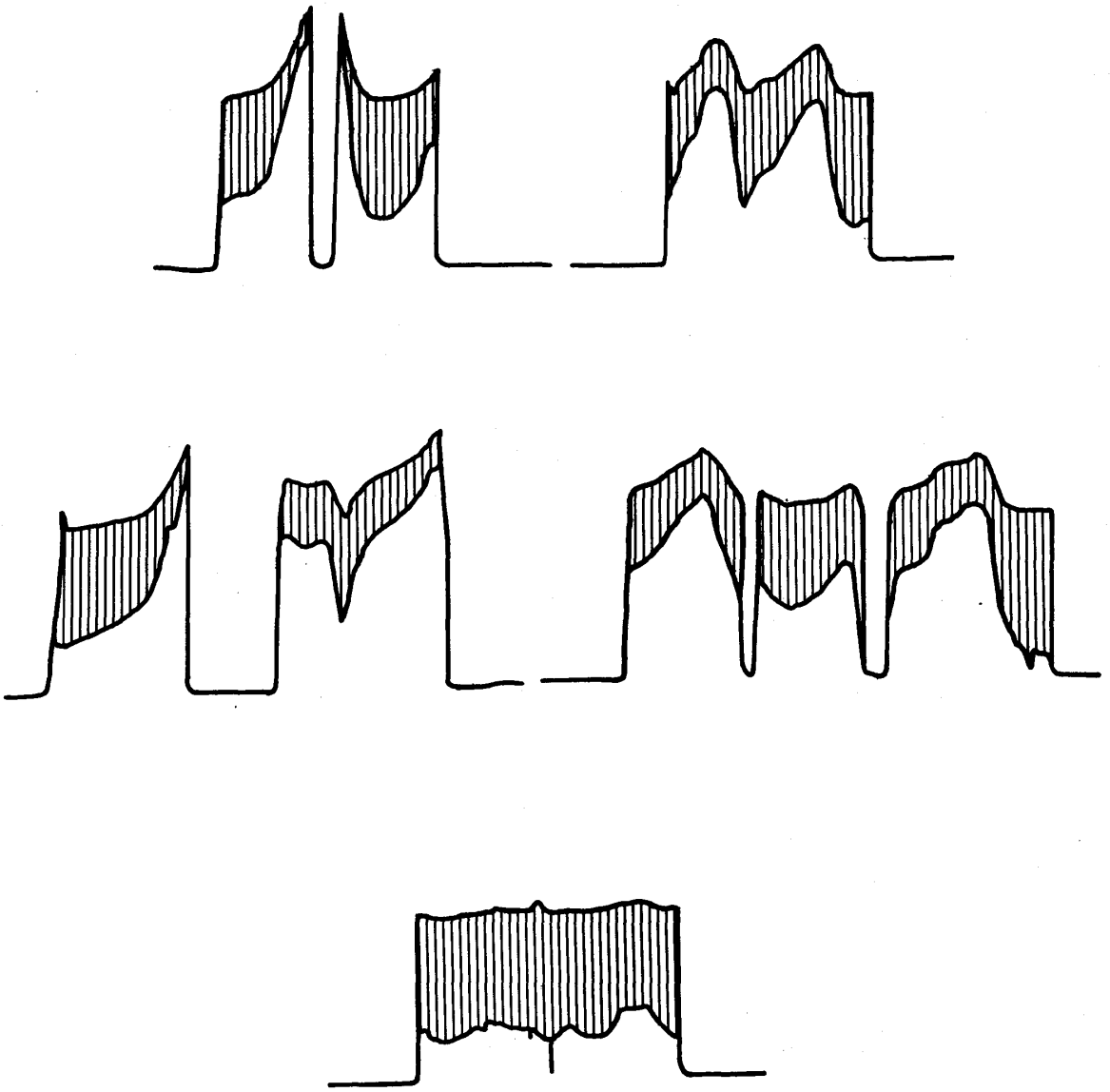

Figure 2. Five patterns of frequency contour produced by an American speaker to record an utterance, "I don't know", with varying extra messages or connotations

Messages Conveyed by Speech Melodies

Variations in speech melody are produced, as it has been observed above, for a variety of reasons. Most convey specific "messages" by specific patterns of melody. Probably one of the most explicit messages expressed and conveyed by the melody is a speaker's emotional condition. When he is "happy", his voice will be high in pitch and the range of pitch change will be wide. On the other hand, if he is "disappointed" or "gloomy" he may speak with a lower pitch, relatively monotonous, or with conspicuously frequent falling inflections. If the speaker is "angry" a higher pitch will be noted. These phenomena had been observed more than 40 years ago in some of the earliest scientific examinations of pitch. Fairbanks and Pronovost (1939) conducted an experiment to compare the characteristics of the voice during the simulation of

emotion. The utterances simulating the emotions of contempt, anger, fear, grief, and indifference were analyzed instrumentally.

They concluded that measureable pitch characteristics distinguish emotional portrayals. Frequency distributions of pitches used in simulating the five emotions are presented in Figure 3.

Further, they summarized the features, as revealed by measurement, which appear to distinguish the five emotions. For instance, in the utterances which were simulating the emotion of CONTEMPT, they found the following features:

1) Extremely wide inflections at the ends of phrases.
2) A low median pitch level in combination with a wide total pitch range.
3) The slowest mean rate of pitch change in narrow inflections.
4) The greatest proportion of upward shifts and the smallest proportion of downward shifts within phrases.
5) The smallest number per second during phonation of pitch changes of all extents.
6) The smallest number per second during the total speaking time of pitch changes less than one semi-tone in extent.

While, in the utterances which were simulating ANGER, they found such features as:

1) The widest mean phonational range.
2) The widest mean extent of inflections in all types and extents of inflections,

⋮

etc.,

(Fairbanks and Pronovost, 1939)

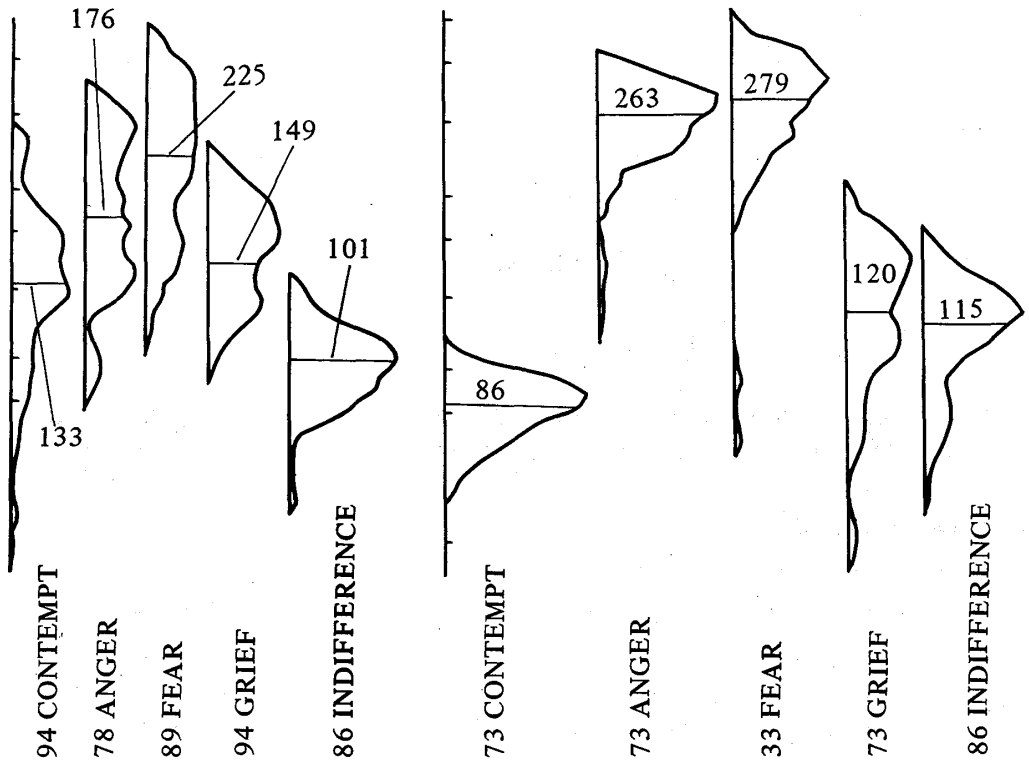

Figure 3. Frequency distributions of pitches used in simulating the five emotions by the two subjects who were most successful (left group) and least successful (right group) in producing identifiable simulations. (from Fairbanks and Pronovost, 1939)

In a more recent examination of emotional messages Hutter (1968) noted that degrees of perceived emotion were highly and positively correlated with fundamental frequency range. He studied the emotional messages ranging in degree on the following scales: bored – interested, excited – calm, passive – active, strong – weak; and in type on the following scales: afraid – bold, confident – timid, sure – unsure, angry – pleased, and happy – sad.

A closely related type of message which can be perceived from the speech melody is that of the speaker's attitude toward the listener(s). It can be judged easily whether a speaker is "polite" to the listener or is "arrogant". He may be even "sarcastic". He may not convey the attitude of "interested" but he may be perceived as trying to be "kind" to the listener. Uldall conducted an experiment using synthesized intonation contours to study the patterns which may be responsible for several attitudinal meanings conveyed in utterances. She chose the following adjective pairs (extremes) to represent three principal attitudes conveyed in intonation: (1) amount or strength of feeling or interest, i.e. "interest", (2) pleasantness or unpleasantness of personal relations, (3) a "power" relationship between speaker and listener, e.g. authority versus submission.

bored . . . . . interested
polite . . . . . rude
timid . . . . . confident
sincere . . . . . insincere
tense . . . . . relaxed
disapproving . . . . . approving
deferential . . . . . arrogant
emphatic . . . . . unemphatic
agreeable . . . . . disagreeable
(Uldall, 1966)

In summarizing the relationships among attitudes, sentence types, and intonation characteristics she observed that authoritative attitude was perceived by American listeners from the patterns of wider ranges and final fall. In contrast, the submissive attitude was perceived in sentences exemplified by narrower range, perturbed weak syllables and final rise (Uldall, 1966). She concluded with the suggestion that the various contours would be differently interpreted by different speech communities. This is in contrast with the relationship of fundamental frequency characteristics and emotional messages which is probably more or less universal (Hutter, 1968). A physiological basis for the changes in pitch functioning which appears to be a concomitant of emotion is the increase in laryngeal tension and muscular activity along with the increase in muscular tension throughout the body (Lindsley, 1951).

Takefuta (1967) found in one of his experiments that most of the foreign accent in English spoken by Japanese lies in pitch characteristics. When he investigated the characteristics of foreign accent in Japanese English psychophysically he found that unnaturalness in melody, rhythm and articulation were significantly correlated with native listeners' perception of foreign accent. However, among the three aspects, relative unnaturalness in melody correlated most highly with the evaluation of foreign accent. More specifically, narrower pitch range and longer periods of nonvocalization were found among the speakers judged to have more foreign accent. Figure 4 presents the speech melody used by Japanese speakers of English in pronouncing a sentence: "We can go alone if you are in a hurry." Speaker 4 (Top left) was rated by native listeners to have the least amount of foreign accent, while speaker 2 (Bottom right) was rated to have the most amount of foreign accent in his utterances.

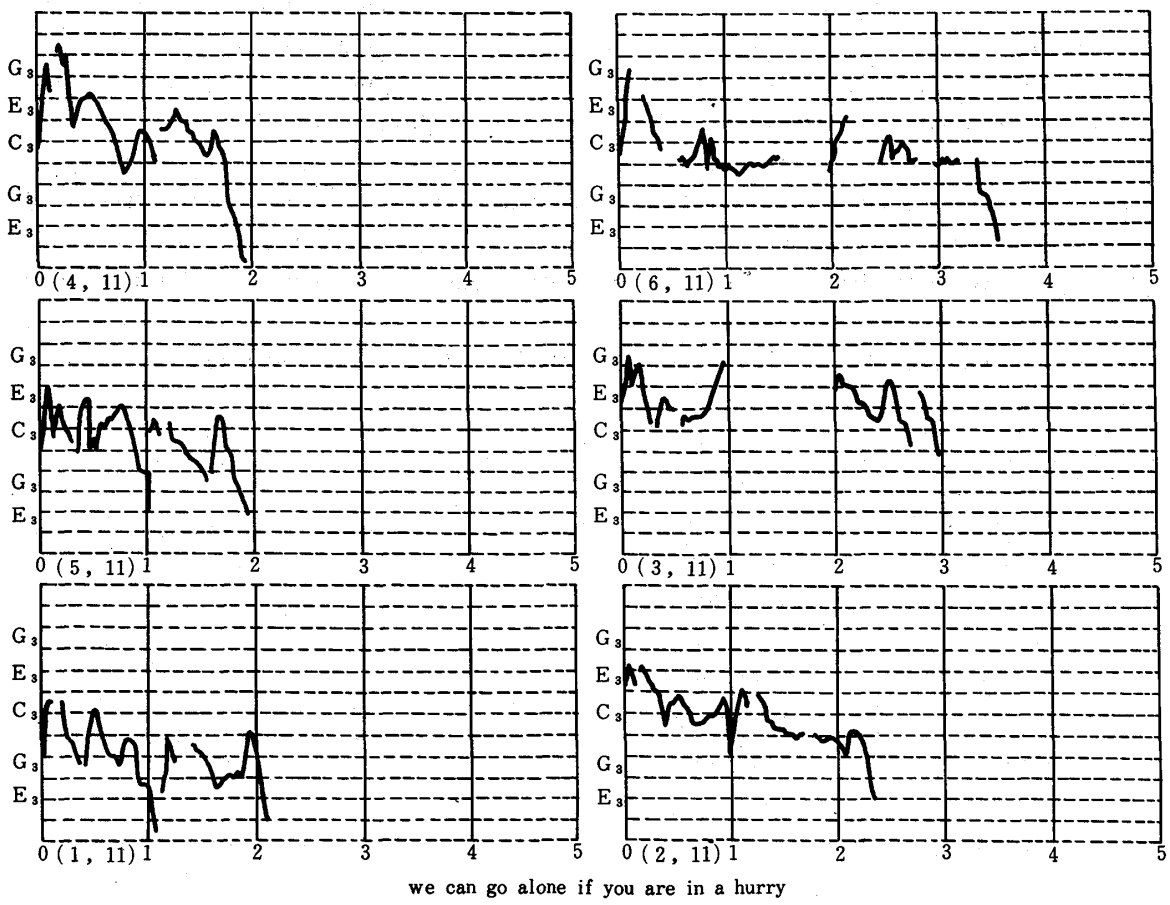

Figure 4. Melody curves produced by six Japanese speakers of English to record a sentence, "We can go alone if you are in a hurry". Speaker 4 (top left) was rated by native speakers to have had the least amount of foreign accent, and speaker 2 (bottom right) to have had the most.

Wise indicated that regional or local dialects in the United States – General American, Eastern, and Southern American English – are also characterized by the difference in intonation which are defined by Hockett as "distinctively different features of speech melody."

> General American English is distinguishable from the speech of other regions by certain characteristics of intonation and pronunciation . . . The speech of the South, like any other regional speech, is characterized by many subtle nuances of duration, pitch, tempo, and rhythm, which defy easy symbolization (Wise, 1957).

The following statement of Wise's, however, is very disappointing, "Despite valuable work on intonation by many authors, from John Hart, 1569, to Kenneth L. Pike, 1945, no simple and highly efficient system of symbolization or notation for intonation has ever been designed."

Fujisaki, et al. (1974, 1976) also analyzed the fundamental frequency contour of two-mora Japanese words in the Tokyo and Osaka dialects and confirmed the relation between its pattern and accent type.

Of major clinical interest is the "extra" messages conveyed by intonation patterns regarding the physiological condition of the speaker. On the simplest level, one can easily detect whether the speaker is "tired" or "fresh and energetic" by listening to him talk. The extreme physiological state would be detection of pathology from pitch characteristics. On the perceptual level of pathologies, West et al. (1937) described a vocal defect:

> Voices lacking clearness of tone may be hoarse, husky, or stridorous. The terms high, shrill, eunuchoid, or treble usually refer to voices lacking pitch levels appropriate to the age and sex of the person being studied. Voices that lack vibrato are said to be hard, metallic, or flat; those that exhibit too much or an irregular or uncontrolled vibrato are described as tremorous and palsied. In terms of inflection or its lack, a voice may show exaggerated pitch change or constantly recurring inflection patterns, such as falling inflection, suggestive of fatigue, at the end of every phrase; or again it may have very little change of pitch and force – in other words, it is monotonous.

Patterns such as these described by West could be the result of neuropathologies, medical problems, hearing loss, emotional disorders, improper vocal habits, delayed sexual development, imitation or onset of puberty.

Highly significant relationships between instrumentally observed frequency characteristics and voice disorders have been experimentally confirmed by many researchers. Weinberg, et al. (1975) found that acromegaly is accompanied by lowering of voice fundamental frequency (FO) characteristics in some patients. Murry (1978) found that speaking fundamental frequency (SFF), standard deviation and semitone range of SFF are significantly reduced for patients with vocal fold paralysis. Spectrographic analysis conducted by Cooper (1974) indicated that 150 of 155 disphonic patients who were speaking with varying degrees of hoarseness were using too low a pitch. This lowered pitch also tended to limit their dynamic range resulting in a more monotonous voice.

The small cycle-to-cycle frequency perturbation (jitter) in the human voice also attracted the attention of many researchers in the field of speech and hearing science. Some expected that hoarse or rough voice quality may be correlated with this phenomenon. Others considered the analysis may help in the early detection of laryngeal pathology. Horii (1979) believes that pathological laryngeal conditions will produce different degrees of FO perturbation for various vowels possibly because of complex muscle interconnections between the larynx and supraglottic articulators.

"Frequency perturbations" can be observed as fundamental frequency (FO) variations around the mean FO of phonation. Davis (1976) observed "pitch perturbation quotients" expressed as FO standard deviation in Hz divided by mean FO. Such frequency perturbation measures other than jitter sizes among consecutive periods as "period relations among every other cycle", "signs of jitter", and the "number of changes from positive jitter to negative jitter, or vice versa" have also been investigated in relation to possible voice defects.

Stereotypical pitch characteristics of utterances may tell the listener the vocation or profession of the speaker. Some ministers and teachers typically show a certain characteristics of lecturing tone in conversation. Other important paralinguistic messages conveyed by pitch

characteristics include the speaker's sex, age and individual personalities. More than a handful of researchers, especially those who had access to the phonelescopic instrument of the University of Iowa studied the pitch and duration characteristics of children, adolescents and aged persons as well as male and female speakers.

Finally, but certainly not insignificantly, messages conveyed by pitch characteristics of utterances are the grammatical information. This type of message is treated last because it departs from the main focus of this work which is the paralinguistic phenomena of language characteristics. Sentence accent or prominence should not be overlooked in this section enumerating messages conveyed by speech melody.

Word accent is also expressed by pitch phenomena. Tone languages, as Chinese and Vietnamese, use change in pitch more significantly but languages such as English or Japanese – not typically considered tone languages – also use pitch characteristics as the signal for word accent (Bolinger, 1958; Fujisaki, 1979; Takefuta, 1980).

It is interesting to note that speech melodies can convey such a variety of messages in our daily conversation. However, it is more important to note that when the meanings defined by words or grammatical structures and those defined by intonation are different the latter often overrides the former. Many instances of this phenomena can be observed easily in our daily conversation. One of the most frequent examples is that of uttering the word "no" with a rapid falling intonation such that the effect is that of incredulity or a complete negation of the meaning of no.

REFERENCES

Berendt, E. A. (1981): "A functional study of teenage English speech," *Eigo no Kenkyu to Kyoiku,* Kirihara Shoten, Tokyo, 173-184.

Berendt, E. A. (1981): "Some aspects of language function in dyadic discourse: clarification and relationship bonding," *Descriptive and Applied Linguistics,* Vol. XIV, (to appear).

Black, J. W. (1942): "A study of voice merit," *Quarterly Journal of Speech,* 28, 67-74.

Black, J. W. (1949): "Natural frequency, duration and intensity of vowels in reading," *Journal of Speech and Hearing Disorders,* 15, 207-210.

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

Bowler, N. W. (1964): "A fundamental frequency analysis of harsh voice quality," *Speech Monographs,* 31, 128-134.

Cooper, M. (1974): "Spectrographic analysis of fundamental frequency and hoarseness before and after vocal rehabilitation," *Journal of Speech and Hearing Disorders,* 39, 3, 286-297.

Cowan, M. (1936): "Pitch and intensity characteristics of stage speech," *Archives of Speech,* Supplement 1, 1-90.

Crystal, T. H., Montgomery, W. W., Jackson, C. L., and Johnson, N. (1970): "Methodology and results on laryngeal disorder detection through speech analysis," Final report, Contract PH-86-68-192, Singnatron, Inc., Lexington, Massachusetts.

Curry, E. T. (1940): "The pitch characteristics of the adolescent male voice," *Speech Monographs,* 7, 48-62.

Davis, S. B. (1976): "Computer evaluation of laryngeal pathology based on inverse filtering of speech," *SCRL Monograph* No. 13, Speech Communications Research Laboratory, Inc.

Denes, P. B. (1970): "The use of speech analysis and synthesis in speech training," *Prosodic Feature Analysis,* Montreal: Librairie Didier, 193-201.

Dewey, G. (1923): *The relative frequency of English speech sounds,* Harvard University Press: Cambridge, Massachusetts.

Fairbanks, G., and Pronovost, W. (1939): "An experimental study of the pitch characteristics of the voice during the expression of emotion," *Speech Monographs,* 6, 87-104.

Fant, G. (1968): "Keynote address," *IEEE Transactions on Audio and Electroacoustics,* AV-16, 1, 3-5.

Fry, D. B. (1968): "Prosodic phenomena," *Manual of Phonetics* (Malmberg, B. ed.), Amsterdam: North-Holland Publishing Co., 365-410.

Fujisaki, H. (1960): "Automatic extraction of fundamental period of speech by autocorrelation analysis and peak detection," *Journal of Acoustical Society of America,* 32, 1518.

Fujisaki, H., Mitsui, Y., and Sugito, M. (1974): "Analysis, synthesis and perception of accent types," *Transactions of the Committee on Speech Research,* Acoustical Society of Japan, s73-51.

Fujisaki, H., and Sugito, M. (1976): "Acoustic and perceptual analysis of two-mora word accent types in the Osaka dialect," *Annual Bulletin of Research Institute of Logopaedics and Phoniatrics,* 10, 157-171.

Gold, B. and Rabiner, L. R. (1969): "Parallel processing techniques for estimating pitch periods of speech in the time domain," *Journal of Acoustical Society of America,* 46, 442-448.

Greenburg, S. R. (1969): "An experimental study of certain intonation contrasts in American English," *UCLA Working Papers in Phonetics,* 13.

Hadding-Koch, K. (1961): *Acoustico-phonetic studies in the intonation of Southern Swedish,* (Lund, Gleerup).

Hecker, M., and Kreul, E. J. (1971): "Descriptions of the speech of patients with cancer of the vocal folds, Part I: Measures of fundamental frequency," *Journal of Acoustical Society of America,* 49, 1275-1282.

Hollien, H., Michel, J., and Doherty, E. T. (1973): "A method for analyzing vocal jitter in sustained phonation," *Journal of Phonetics,* 1, 85-91.

Horii, Y. (1979): "Fundamental frequency perturbation observed in sustained phonation," *Journal of Speech and Hearing Research,* 22, 1, 5-19.

Hutter, G. L. (1968): "Relations between prosodic variables and emotions in normal American English utterances," *Journal of Speech and Hearing Research,* 11, 481-487.

Lehiste, I., and Peterson, G. E. (1961): "Some basic onsiderations in the analysis of intonation," *Journal of Acoustical Society of America,* 33, 419-425.

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

Lieberman, P. (1972): *Speech acoustics and perception,* Indianapolis: The Bobbs Merrill Company, Inc.

Lindsley, D. B. (1951): "Emotion," in *Handbook of experimental psychology* (Stevens, S. S. ed.), New York: John Wiley.

Linke, C. E. (1953): "A study of pitch characteristics of female voices and their relationship to vocal effectiveness," Doctoral dissertation, State University of Iowa.

Lynch, G. E. (1936): "A phonophotographic study of trained and untrained voices reading factual and dramatic material," *Archives of Speech,* 1, 9-25.

Meyer-Eppler, W. (1957): "Realization of prosodic features in whispered speech," *Journal of Acoustical Society of America,* 29, 104-106.

Murry, T. (1978): "Speaking fundamental frequency characteristics associated with voice pathologies," *Journal of Speech and Hearing Disorders,* 43, 3, 374-378.

Noll, A. M. (1967): "Cepstrum pitch determination," *Journal of Acoustical Society of America,* 41, 293-309.

Roberts, A. H. (1965): *A statistical linguistic analysis of American English,* The Hague: Mouton.

Snidecor, J. C. (1943): "A comparative study of the pitch and duration characteristics of impromptu speaking and oral reading," *Speech Monographs,* 10, 50-56.

Takebayashi, S. (1970): "A comparative study of Japanese English consonant phonemes," *Essays in Honor of Claude M. Wise* (Bronstein, A. J., Shaner, C. L., and Stevens Cj, eds.) Columbia: Missouri, Artcraft Press, 97-114.

Takefuta, Y. (1964): "The pitch and duration of English spoken by Japanese," Unpublished Master's thesis, The Ohio State University, Columbus.

Takefuta, Y. (1967): "A study of foreign accent in Japanese English," *Speech Monographs,* XXXIV, 167-171.

Takefuta, Y. (1975): "Method of acoustic analysis of intonation," *Measurement Procedures in Speech, Hearing, and Language* (Singh, ed.), Baltimore: University Park Press, 363-378.

Takefuta, Y., and Jancosek, E. G. (1976): "Perception, recognition, and analysis of American intonation by human listeners and by a computer," *Study of Sounds,* Tokyo: Phonetic Society of Japan, 89-108.

Takefuta, Y., and Suzuki, S. (1980): "Extraction of the patterns of English stress and Japanese word accent by digital data processing," *Descriptive and Applied Linguistics,* Tokyo, Japan: International Christian University, 153-167.

Uldall, E. T. (1960): "Attitudinal meanings conveyed by intonation contours," *Language and Speech,* 3, 223-233.

Van Riper, C. (1939): *Speech correction principles and methods,* Englewood Cliffs, N. J.: Prentice-Hall.

Voelker, C. H. (1942): "The one-thousand most frequent spoken-words," *Quaterly Journal of Speech,* XXVIII, 189-197.

Weinberg, B., Dexter, R., and Horii, Y. (1975): "Selected speech and fundamental frequency characteristics of patients with acromegaly," *Journal of Speech and Hearing Disorders,* 40, 253-259.

West, R., Kennedy, L., and Carr, A. (1937): *The rehabilitation of speech,* New York: Harper and Brothers.

Wise, C. M. (1957): *Applied phonetics,* Englewood Cliffs, N. J.: Prentice-Hall.