GRAPHICAL REPRESENTATION OF PERCEIVED PITCH IN SPEECH

J M. COWAN

There are at least three different kinds of *pitch* in speech, defined by the manner in which they are isolated. The earliest of these to receive attention we shall call *impressionistic.* Analysts of impressionistic pitch hear pitch differences, levels and fluctuations, and attempt to sketch these graphically on a relative scale of high and low with temporal sequence from left to right.[1] Several systems of such notations have evolved.

In an attempt to obtain more "accurate" representations of pitch, instrumentalists have isolated the *physical* pitch, that is, the average frequency per time unit of the vibrations of the vocal cords. There has been a marked development in elegance of the instrumentation for this purpose, but no appreciable change in the graphical representation of the results.[2] The plot has pitch on the ordinate, usually logarithmic, and time on the abscissa, only the aspect ratio varies from report to report.

The third kind of pitch we shall call *linguistic.* Within the framework of modern descriptive linguistics, only those units which constitute significant contrasts within a given language are taken into consideration. Roughly, these are of two kinds, pitch phonemes which are tone levels or contours centering on a vowel nucleus, and intonation morphemes signalling meaning contrasts over longer utterances. The linguist has to date been forced to rely primarily on impressionistic pitch analysis in his work because of tediousness and complexity of instrumental analysis and certain difficulties inherent in attempts to evaluate the raw results of instrumentation.[3]

There has been a naive tendency on the part of instrumentalists to assume that

[1] Sweet, H., *Primer of Spoken English* (Oxford, 1906); Klinghardt, H., *Uebungen im deutschen Tonfall* (Leipzig, 1927); Armstrong, L. E., and Ward, I. C., *Handbook of English Intonation* (Leipzig, 1926); Schubiger, Maria, *English Intonation* (Tübingen, 1958).

[2] Scripture, E. W., *Researches in Experimental Phonetics* (Washington, D.C., 1906); Cowan, J M., "Pitch and Intensity Characteristics of Stage Speech," *Archives of Speech*, Supplement (Iowa City, 1936); *id.*, "A Technique for the Measurement of Intonation," *Arch. f. vergl. Phonetik*, 3:223–234 (1939); Grützmacher, M., and Lottermoser, W., "Die Verwendung des Tonhöhenschreibers bei mathematischen, phonetischen und musikalischen Aufgaben," *Akust. Zeitschrift*, 3:183–196 (1938); Lehiste, I., and Peterson, G. E., "Some Basic Considerations in the Analysis of Intonation," *Jour. Acous. Soc. Amer.*, 33:419–425 (1961); Mathews, M. V., Miller, Joan E., and David, E. E., Jr., "Pitch Synchronous Analysis of Voiced Sounds," *Jour. Acous. Soc. Amer.*, 33:179–186 (1961).

[3] Pike, Kenneth, *Intonation of American English* (Ann Arbor, 1945); Smith, H. L., and Trager, G. L., *An Outline of English Structure* (Norman, Oklahoma, 1951).

their graphical records of the cord tone were fairly adequate representations of the perceptual pitches in the speech analyzed. In 1936 the author sounded the following warning. "The graphical speech scores in the text are accurate representations of physical characteristics of continuous discourse. As such they throw light upon the physiological function of the speech mechanism. In some instances they also give a fairly adequate representation of the perceptual aspects of speech. Certainly they are the closest approximation to such representation known to us at the present time. However, the analyses have brought out the fact that there is frequently noticeable discrepancy between the physical stimulus and the anticipated corresponding perception. Any attempt to evaluate the scores perceptually must depend on further psychophysical work." A modest amount of psychophysical work has since been done in an attempt to derive transformation rules which would permit perceptual evaluation of physical pitch graphs.[4] However, the problems seem to mount faster than the solutions.

The three kinds of pitch have by no means been kept neatly separated in the various studies of speech pitch. Even simple impressionistic representations have had an overlay of what is called stress or accent, the latter not always carefully defined in linguistic terms. Pike related impressionistic and linguistic pitch (see footnote 3); Bolinger explored physical and linguistic pitch.[5] Lewis, Cowan and Fairbanks and Zwirner combined physical and impressionistic observations (= psychophysical) (see footnote 4).

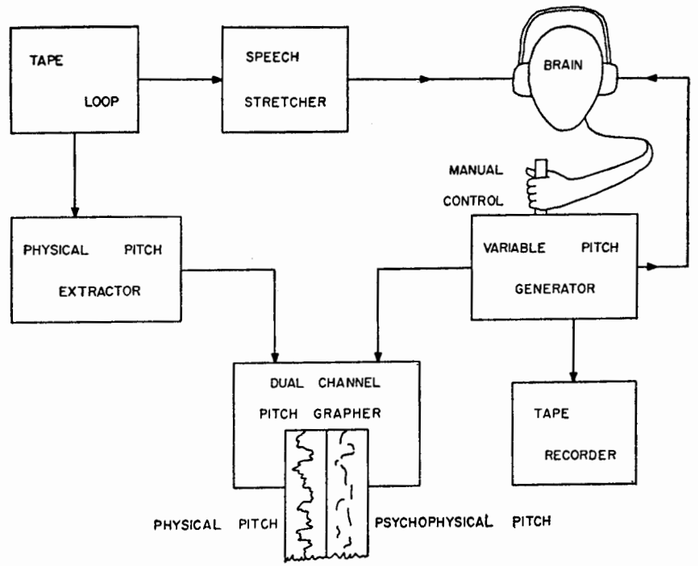
Lewis[6] has treated pitch as a psychological phenomenon, as a basic, scalable attribute of sound and discussed its physical determinants. It seems logical, for all purposes, to isolate perceived (psychophysically determined) pitch as a first step and relate it subsequently to any other factors, such as linguistic data. This paper describes a method of doing this.

The physical determinants affecting pitch perception in speech are numerous and highly variable in time. Vocal pitch, from the linguistic point of view, is the carrier which bears an overlaid succession of significant phones occurring at an average rate (in most European languages) of 10–20 per second. These latter place such a load on the attention of the listener that it is very difficult to abstract the underlying intonation from the rest of the sound complex. The isolation of pitch can be accomplished using a matching technique.

[4] Peters, W. E., *Die Auffassung der Sprechmelodie* (Leipzig, 1924); Root, A. R., "Pitch-Patterns and Tonal Movement in Speech," *Psychological Monographs, Univ. of Iowa Studies in Psychol.*, No. 13:109–159 (1930); Zwirner, E., "A Contribution to the Theory of Pitch Curves," *Archives Néerl. de Phon. Exp.*, 7:1–14 (1932); Zwirner, E., and Zwirner, K., "Ueber Hören und Messen der Sprachmelodie," *Archiv f. vergl. Phonetik*, 1:35-47 (1937); Lewis, D., Cowan, J M., and Fairbanks, G., "Pitch and Frequency Modulation", *Jour. of Exp. Psychology*, 27:23–36 (1940).

[5] Bolinger, D. L., "A Theory of Pitch Accent in English", *Word*, 14:109–149 (1958).

[6] Lewis, Don, "Pitch: Its Definition and Physical Determinants," *Univ. of Iowa Stud. Psycho. Music*, 4:346–373 (1937); *id.*, "Pitch as a Psychological Phenomenon," *Proc. Music Teachers National Assocn.*, 121–133 (1939).



The figure shows a schematic representation of the instrumentation. The speech sample to be analyzed is recorded on a tape loop for convenient repetition. This signal is fed into one ear of the analyst. In order to lighten the task, the speech sample is expanded 100 per cent in time by means of the stretcher. A matching tone of approximately the harmonic structure of the human cord tone is fed into the other ear. This matching tone is continuously variable in frequency and is adjusted by one hand of the analyst. His other hand operates an on-off micro-switch, enabling him to match the presence or absence of vocal pitch. After a brief training period an operator can learn to match accurately the perceived vocal pitch progression. When a successful match has been made a psychophysical experiment has been completed and successive matches of the same speech signal by different operators will show a high order of convergence.

The matching tone is recorded continuously on a logarithmic graph and also on a tape recorder. The visual image of a successful pitch match can thus be compared both with the original signal and with an auditory playback of the matched pitch minus the overlaid modulations of speech sounds.

A convenient, but unnecessary, elaboration of instrumentation is indicated by the inclusion of simultaneous graphing on the second recording channel of the physical pitch of the speech sample as analyzed by an electronic glottal-frequency extractor. A comparison of the two graphs quickly points up the fallacy of considering the physical pitch to be an adequate representation of the perceived (psychophysically

determined) pitch. However, extensive study of such two channel graphs should be the most convenient manner of arriving at the complex psychophysical relations existing between stimulus and perception in this field.

An important modification of the instrumentation as described, involves feeding the output of the manually controlled pitch generator into an artificial larynx and having the operator reproduce the speech sample without using his vocal cords. We have used a set-up similar to that described by Firestone.[7] Many operators feel more confident of their pitch matching during their initial training on the equipment when they are using the artificial larynx. All should submit to this training in order to sharpen their attention focus and manual dexterity with the pitch generator. The essential difference in using the larynx is that the operator listens first to the signal and then attempts to reproduce it. In straight pitch matching, the matching takes place simultaneously with receipt of the original signal in the brain.

Three areas of application are apparent. The possibilities for the psychophysics of audition have already been mentioned. Descriptive linguistic analysis will be benefited. Linguists have become increasingly aware of the importance of intonation in their field but have been hindered from giving it systematic treatment by the complexity and plurality of factors affecting it. It should now be possible to proceed rapidly from the graphical and auditory records derived from this kind of equipment and procedure to systematic statements about the linguistics of intonation (and/or tonemics) of any given language. It should be mentioned at this point that for English it has been found quite simple to "punch out" syllabic pitch levels and terminal contours. This ability is different from that of simply matching the perceived pitch progression. But it can be quickly acquired by a trained operator. Finally, there are promising pedagogical applications, especially in foreign language instruction. The graphical records and accompanying playback of the pitch patterns can illustrate vividly to language learners the differences between their native intonation patterns and those of the target language. It is common knowledge that these native patterns are so deeply ingrained (before the age of six) that a learner of a foreign language may imitate very closely the individual phones of a foreign language and speak it almost unintelligibly or with a very heavy accent simply because of the carry-over from his own intonation patterns.

Cornell University

[7] Firestone, F. A., "An Artificial Larynx for Speaking and Choral Singing by One Person," *Jour. Acous. Soc. Amer.*, 11:357–361 (1940).