Working Papers 12 · 1975
Lund University

Phonetics Laboratory
General Linguistics

FUNDAMENTAL FREQUENCY OF THE VOICE IN CONTINUOUS SPEECH. – PRELIMINARY REPORT ON A DEVICE FOR DETERMINING MEAN AND DISTRIBUTION OF FREQUENCIES.

Peter Kitzing*, Hans-Erik Rundqvist** and Ewa Talo*

Abstract

There is no generally accepted and practicable method for quantifying average vocal pitch during continuous speech, even though this parameter is very often affected in different kinds of dysphonias and is the object of therapeutic measures in clinical work with voice disorders.

The present paper is a description of an apparatus for immediate presentation of the mean frequency and the distribution of frequencies of the fundamental of the voice during continuous speech. Some frequency distributions for normal and pathological voices illustrate the function of the apparatus.

There are various approaches to the study of the fundamental of the voice during continuous speech. Phoneticians are mostly interested in the linguistic implications of variations of pitch or intonation. On the other hand, the clinical interest of speech pathologists and phoniatricians is more often directed to vocal pitch as a correlate of laryngeal function. From this latter point of view it is of greater interest to obtain objective data about a person's total intonation

---

* Dept of Phoniatrics, ENT Clinic, University Hospital, Malmö.
** Dept of Electrical Measurements, Institute of Technology, Lund.
Address of corresponding author:
P. Kitzing, M.D.
Talvårdsavdelningen
Allmänna Sjukhuset
S-214 01 MALMÖ
Sweden

range and the mean and distribution of the fundamental frequencies of the voice during continuous speech rather than about language-dependent intonation patterns.

For phonetic research, many methods have been developed for extracting the fundamental of the voice from isolated sounds or continuous speech. Most of these methods consist of two steps. First, the vibratory frequency of the vocal folds or the corresponding fundamental frequency of the voice is recorded in some way. Then the recorded data are measured, which is usually a laborious and time-consuming procedure. This is the case for high-speed cinematography, for glottograms, for oscillograms and for sound spectrograms.

One very common method is to use a frequency meter to produce an intonation curve. Such meters have been described by Grützmacher and Lottermoser (1937), Risberg (1962) and Svend Smith (1968), all having in common a low-pass filter in their initial step. The bandwidth of this filter must be selected individually for different speakers, in order to pass the highest occurring fundamental frequencies while at the same time excluding the harmonics. This is not always feasible, so that measurements with this kind of frequency meter are liable to be somewhat uncertain (Fry, 1974). Moreover, the manual calculation of the mean fundamental frequency and the frequency distribution from an intonation curve is very time consuming.

Another possibility is to use a digital computer to extract the fundamental frequency of the voice. Several different computer programs have been reported recently (e.g. Maksym, 1973).

In the field of speech pathology and voice disorders it has not been possible to use either the laborious or the sophisticated methods mentioned above for routine clinical work. Instead, clinicians have checked the performance of their patients with aid of tuning forks or some musical instrument (e.g. Böhme and Hecker, 1970). While it seems quite possible to obtain a valuable estimate of the voice range in this way, the method is obviously far too subjective to offer a reasonably safe estimate of the average fundamental pitch of the voice in continuous speech. On the other hand, the average fundamental pitch in continuous speech is one of the most important aspects of vocal function, and voice therapy often aims at normalizing this parameter when it is pathologically disturbed.

The present paper is a description of a specially designed apparatus for on-line determination of the mean and distribution of fundamental frequencies of the voice. It is easy to use, and its function will be described and illustrated with some distributions for normal and disturbed voices.

Method

The principal components of the preliminary version of the system appear in the block diagram, Fig. 1. The signal from a microphone (Sennheiser MD 421 N) is amplified and filtered in a low-pass filter (Fonema Type 0031) set to cut off the frequencies above the fundamental. The resulting signal is converted by a Schmitt trigger circuit into rectangular wave form. The periods are measured in a digital period meter and the frequency calculated as the reciprocal of the period.

The rectangular wave is passed to the mean fundamental frequency meter (with digital display) and to the frequency distribution processor (the results of which are displayed as a histogram on an oscilloscope screen).

In the mean fundamental frequency meter the number of cycles is counted during a phonatory time of twenty seconds. As voiced and unvoiced speech sounds alternate in continuous speech, it is essential that the device operates only when a voiced input signal is present. At the end of the processing the mean fundamental frequency appears on a three digit display.

The frequency distribution processor consists mainly of a number of memory cells, each representing a given frequency interval. In the present preliminary version of the apparatus the number of memory cells has had to be limited to nine. The frequency interval of each cell was chosen to be 6/7 of a whole tone, which means that in its present form it can cover a range of about 16 semitones. This range can be adjusted in six steps as appropriate for different voices by means of a special frequency range selector. Each incoming cycle is assigned to the relevant cell, according to its frequency. The number of cycles counted in each cell is scanned at the end of the processing, and the distribution of the frequencies (strictly, of individual cycle periods) of the fundamental of the voice in the analysed speech sample is displayed as a histogram on the oscilloscope screen. Each histogram column represents the frequency interval of the corresponding memory cell while its height is proportional to the number of cycles counted in that interval. (For technical details cf. Rundqvist, 1974.)

In practice, the mean fundamental frequency can be read off directly from the digital display. But the histograms produced by the present version of the apparatus cannot be used without further processing. In order to compare histograms obtained on different occasions, the number of cycles counted in each cell must first be expressed as a percentage of the total since the absolute numbers can vary from one sample to another.

The speech sample analysed was a paragraph of neutral test ("Nordanvinden och solen", The North Wind and the Sun) comparable to the well known "Rainbow" passage (Fairbanks, 1960) read at the natural intensity of quiet conversation.

Preliminary results

The function of the new apparatus can be illustrated with the following examples.

In normal male (Fig. 2a) and female (Fig. 3) voices there is a fairly wide range of fundamental frequencies extending over six or seven columns in the histogram (5 or 6 tones), not counting columns with height lower than 5 %. The histogram shows either one or two peaks, neither normally exceeding a height of 30 %. Intonation can of course be restricted volontarily and this is illustrated by a sample of artificially monotonous speech with almost 80 per cent of all cycles occurring within the same interval (Fig. 4).

Restriction of the intonation range may be one feature of functional voice disorders as can be seen in Fig. 5, which is the record of a case of phonasthenia before and after voice therapy. Before therapy, the fundamental frequency distribution extended over four columns with two of them exceeding 30 %. After therapy, the average fundamental pitch was lowered about one semitone and the frequency distribution widened by one interval with no column in the histogram exceeding 30 %. Similar results have usually been found in other cases of functional voice disorder before and after therapy.

In organic voice disorders the same type of voice changes may occur. This is illustrated in Fig. 6, recorded from a 32 year old male with a left vocal fold paresis of unknown etiology. His speaking voice sounded quite normal although somewhat weak, but there were typical voice breaks when coughing and laughing. Being a talented amateur singer, he suffered from the loss of his singing voice. But in spontaneous speech he was able to control his voice unusually well and so did not experience great discomfort from his paresis during normal conversation. Therefore, there was no need for voice therapy but, of course, he was re-examined at regular intervals. One year after the appearance of the paresis, laryngolocial examination revealed his left vocal fold had regained normal respiratory and vibratory function, and the patient reported a normalization of his voice when speaking as well as singing or shouting. His mean voice pitch was now perceptibly lower and in his vivid intonation he covered a wide frequency range. The course of events described is objectively preserved in the fundamental frequency record of his speech, where his average fundamental pitch is lowered by two semitones and his frequency range widened from four intervals to six (7-10 semitones), not counting columns containing less than 5 % (Fig. 6).

In the cases of functional and organic dysphonias reported above, and in other frequency recordings not reported here, the amount and nature of voice change has been checked auditively by listening to tape recordings without prior knowledge of the fundamental frequency analysis. The results of this subjective evaluation, which of course could be expressed only in a general description, were in good accordance with the numerical data obtained by the objective frequency analysis.

Finally, the reproducibility of the method is illustrated by the two examples from a healthy male subject shown in Fig. 2, with a time interval of about 20 minutes between recordings.

Discussion

Our original goal was to construct a mean fundamental frequency meter for continuous speech which would easily yield objective data on-line for daily clinical examinations of patients with voice disorders. It became clear during preliminary discussions that a display of the distribution of frequencies would also be of great interest. This opinion was substantiated after searching the literature on the fundamental frequency distribution of speech. This literature is quite sparse, probably because of the hitherto very laborious data-gathering procedure. We are aware of only four papers about fundamental frequency distribution in continuous speech, namely Smith (1955), Risberg (1961), Saito et al. (1958) and Pisani (1971).

According to the highly experienced voice therapists of the State Institute for Speech Disorders in Copenhagen, Smith and Lauritzen (Smith 1955), intonation is always changed in a characteristic way after pedagogical voice therapy for functional voice disorders. As illustrated in Smith's paper, Lauritzen was able to demonstrate these changes in diagrams of the frequency distribution from four patients with functional voice disorders. He obtained his original data on fundamental frequency of the voice by measuring an oscillogram curve at sampling intervals of 20 ms. Two characteristic changes after therapy are reported. Firstly, there is most often an influence on the fundamental frequency of the voice, the directin of which seems to depend on the kind of voice disorder and the "placing" of the voice by the voice therapist. Secondly, there is usually an extension of the fundamental frequency range of the voice, displayed by a "widening of the peak" of the distribution or by "emergence of two peaks" (quotations from Smith).

In a pilot study on the range and rate of change of fundamental frequency in continuous speech, Risberg (1961) published frequency distribution charts from three male and two female speakers reading Swedish or English texts. Fundamental frequency data were sampled manually at successive 25 ms intervals from a Mingograph oscillogram produced by a modified Grützmacher analyzer. The speakers were described as being trained, semitrained or untrained. In the published diagrams there seems to be a tendency for widened frequency range to be positively correlated with the degree of the speaker's training.

Our own preliminary results are in good accordance with the two pilot studies just quoted.

Pisani (1971) reports another modification of the Grützmacher frequency analyser. By the use of a multichannel analyser the statistical distribution of voice fundamental frequency can also be detected. The paper includes uncommented records of fundamental frequency distribution from the same subject reciting two different pieces of poetry.

Another type of voice frequency distribution analysis has recently been reported by Müller (1973). The method is said to be based on the filter bandwidth of the ear and measures the time during which certain trigger levels are reached in 24 frequency bands reanging from 20 to 15 500 Hz. In this way the entire acoustic speech signal is analysed at various trigger levels for amplitude distribution in each of the 24 frequency bands. As the method yields an analysis of speech rather than voice, the resulting records arise from joint influence of at least three parameters, namely voice fundamental frequency, formant regions and filter bandwidth. This complex origin of the records complicates the interpretation of the resulting curves. The analysed signal to a certain degree stems not only from the voice source but also from the resonances of the vocal tract. This will render a correlation of these curves to the subtle auditive criteria of functional voice disorders very difficult.

In all methods of analysis discussed so far, as well as in our own preliminary investigations, the original signal analysed has been acoustical. As mentioned earlier in the introduction, the detection of the voice fundamental in an acoustic speech signal is subject to risks of error, mostly arising from unsufficient separation of harmonics from the fundamental. However, experience from the detection of laryngeal vibrations with the aid of an electroglottograph has convinced us that these difficulties can be overcome. Therefore we have recently substituted a glottograph for the microphone in our fundamental frequency analyser, so that the input signal corresponds to the laryngeal vibrations and is not an acoustic speech signal.

The method just described seems to be a practicable and economic way of obtaining objective data about normal and pathological voice pitch in spontaneous speech. The relevance of these data must of course be further substantiated by correlating the new fundamental frequency analyses to the auditive evaluation of the same voice samples by a panel of experienced listeners. Furthermore, the influence on intonation range of different types of text for speech samples, and the speaker's age, sex and dialect must be investigated. Once these aspects are fully understood, the method is expected to become a useful tool in clinical work with patients suffering from voice disorders.

References

Böhme G. and Hecker G. 1970. Gerontologische Untersuchungen über Stimmumfang und Sprechstimmlage. Folia phoniat. 22, 176-184

Fairbanks G. 1960. Voice and articulation drill book. Harper & Row, New York, Evanston, and London. 2:d ed.

Fry D.B. 1974. Prosodic phenomena. In Malmberg B., Manual of phonetics. North-Holland/American Elsevier, Amsterdam. 3rd printing

Grützmacher M. and Lottermoser W. 1937. Uber ein Verfahren zur trägheitsfreien Aufzeichnung von Melodiekurven. Akust Z. 2, 242-248

Maksym J.S. 1973. Real-time pitch extraction by adaptive prediction of the speech wave form. IEEE Transactions on Audio and Electroacoustics, AU 21:3, 149-154

Müller G. 1973. Statistische Langzeitanalyse von Sprechstimmen auf Frequensgruppenbasis. Arch. Ohr.-, Nas.-, Kehlk.-Heilk., 204 151-160

Pisani R.U. 1971. A melody detector. Acustica 25:3, 179-182

Risberg A. 1961. Statistical studies of fundamental frequency range and rate of change. Speech Transmission Laboratory, Quaterly Progress and Status Report 4, 7-8

Risberg A. 1962. Fundamental frequency tracking. Proc. 4th Int. Congr. Phonet. Sci., 227

Rundqvist H.-E. To be published. Device for registration of mean value and distribution of fundamental frequencies of the human voice in continuous speech

Saito S., Kato. K. and Teranishi N. 1958. Statistical properties of the fundamental frequencies in Japanese speech voices. J. Acoust. Soc. Japan 14, 111-116 (quoted from Risberg, 1961)

Smith S. 1955. Om tonehøjde og maaling af denne. Nord. Tidsskr. Tale og Stemme 15, 57-69

Smith S. 1968. A valve device for difficult cases of F_o recording. Folika phoniat. 20, 202-206

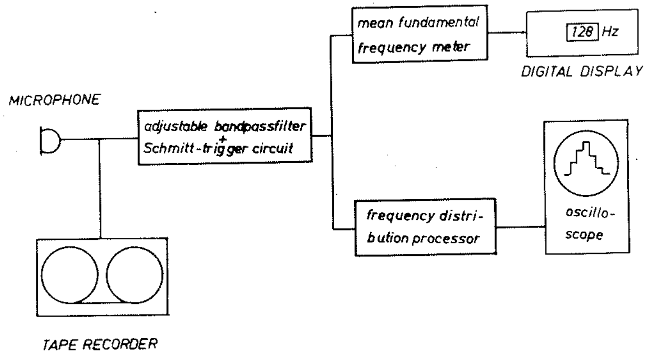

Fig. I

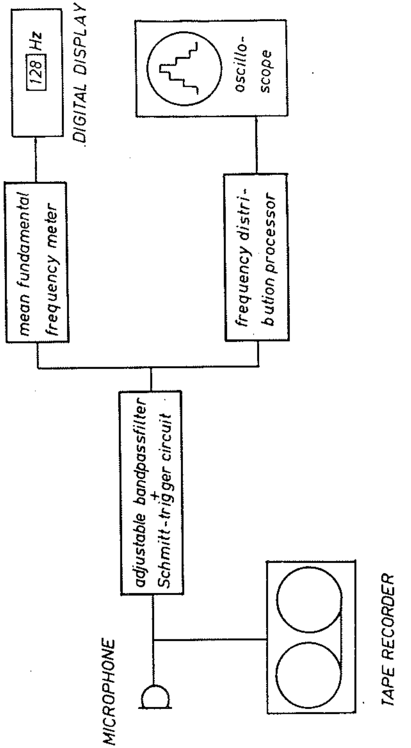

Fig. 1. After low pairs filtering to exclude harmonics, the voice signal is transformed by a schmitt trigger circuit into rectangular waveform with the same cycle period as the corresponding cycle of the voice fundamental. Further processing of the signal yields immediate digital display of the mean fundamental frequency (MFF) of the voice and a histogram of the distribution of frequencies on an oscilloscope.

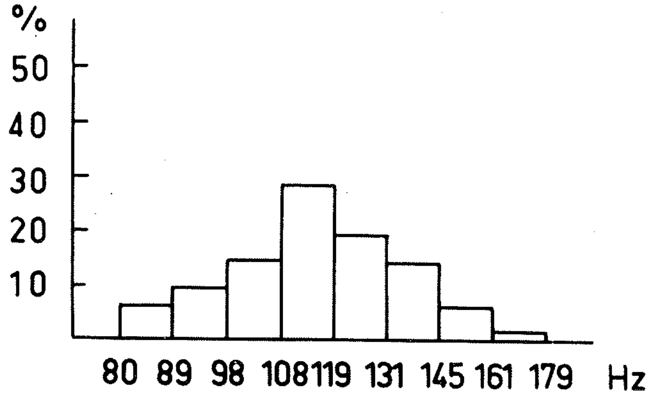

Fig. 2a

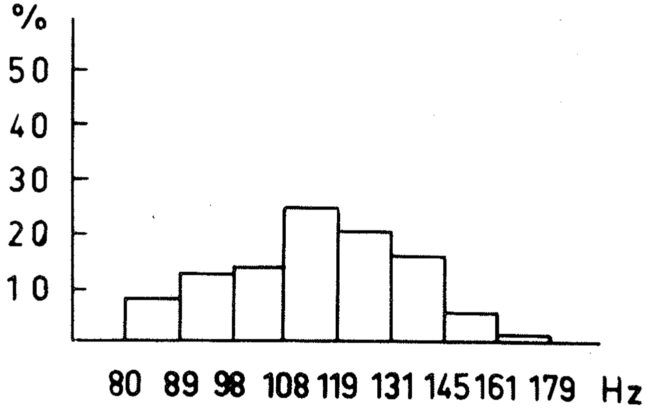

Fig. 2b

Fig. 2. a. Fundamental frequency distribution of a normal voice when reading a paragraph of neutral text in a natural way. The ordinate shows the number of cycles counted in each frequency interval expressed as a percentage of the total number of cycles counted in all cells. Male subject, aged 40. MFF: 112 Hz.

b. Distribution from the same subject 20 minutes later, demonstrating the reproducibility of the method. MFF: 110 Hz.

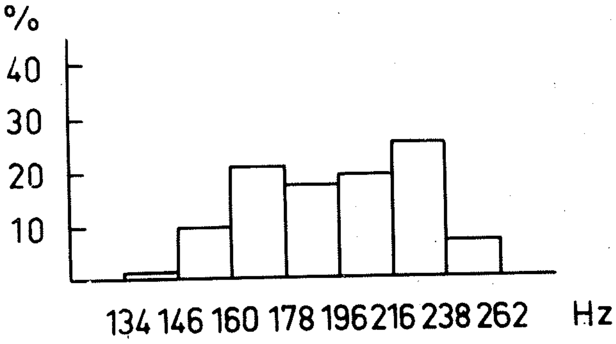

Fig. 3. Fundamental frequency distribution of a normal voice. Female subject, aged 30. MFF: 192 Hz.

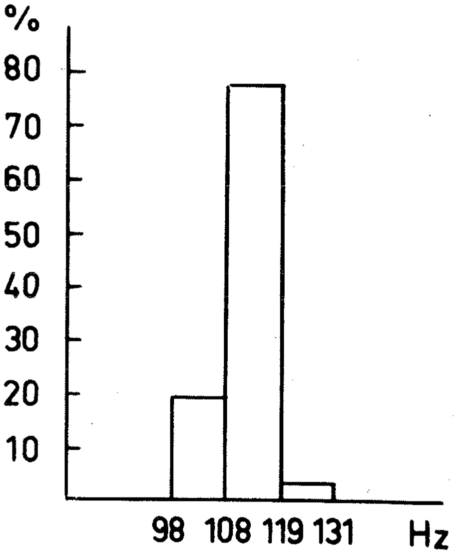

Fig. 4. Fundamental frequency distribution when reading in an artificially monotonous way. Male subject, aged 40. MFF: 109 Hz.

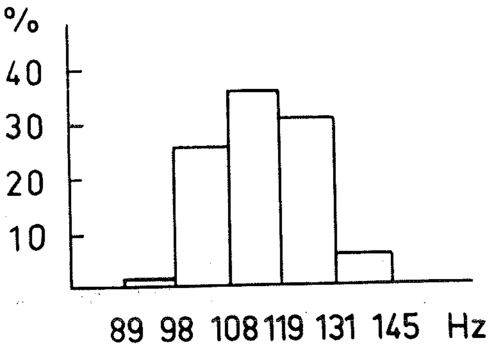

Fig. 5a

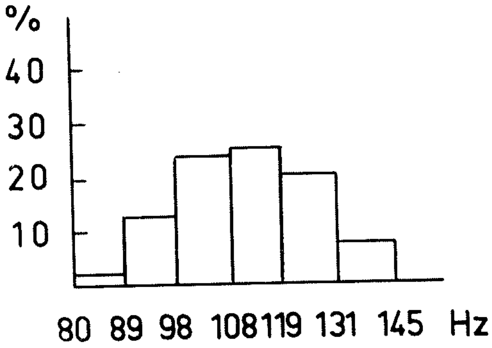

Fig. 5b

Fig. 5. Fundamental frequency distribution during continuous speech for a case of functional voice disorder (phonasthenia)
(a) before therapy. and
(b) after voice therapy and normalization of the voice function. Male subject, aged 32. MFF before therapy: 127 Hz; after therapy: 118 Hz.

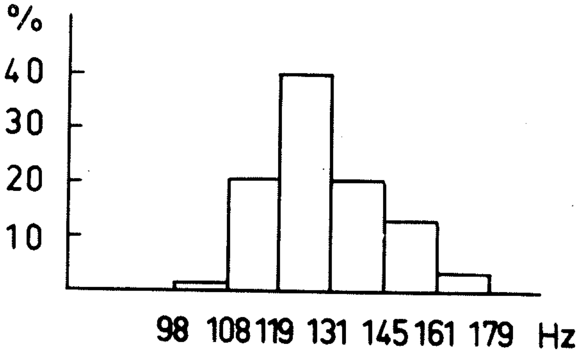

Fig. 6a

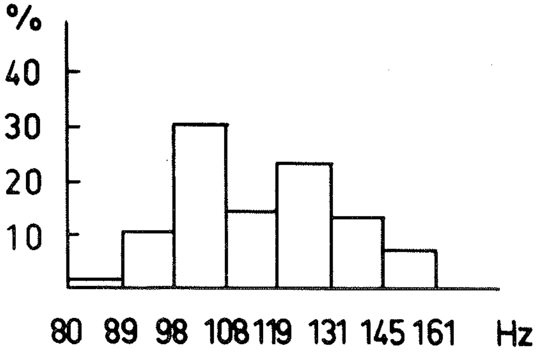

Fig. 6b

Fig. 6. Fundamental frequency distribution during continuous speech for a case of organic voice disorder (vocal fold paresis on one side)
(a) during paresis, MFF: 134 Hz, and
(b) after spontaneous recovery. MFF: 121 Hz. Male subject, aged 32