ARCHIVES OF ACOUSTICS
5, 1, 3-30 (1980)

AN ACOUSTIC METHOD FOR THE EVALUATION OF THE STATE OF THE LARYNX SOURCE IN CASES INVOLVING PATHOLOGICAL CHANGES IN THE VOCAL FOLDS*

RYSZARD GUBRYNOWICZ, WŁADYSŁAW MIKIEL, PIOTR ŻARNECKI

Institute of Fundamental Technological Research, Polish Academy of Sciences
(00-049 Warszawa)

With a view to discussion of the role of the larynx source in the formation of the acoustic speech signal, this paper presents a method for the evaluation of the state of the larynx for different pathological states of the vocal chords. 143 persons, both healthy and with specific pathological voice changes were investigated. The pathological conditions included unilateral and bilateral paralysis of the larynx nerves, inflammation states and cancerous development, both benign (polyps, papillomas etc.) and malignant. For the analysis of the speech signal an Intonograph-77 was used, which together with a MERA-300 minicomputer provides an integral system for automatic measurement and registration of the values of successive periods of the larynx tone, currently determined, during the articulation of longer texts. The evaluation of the state of the larynx source was performed on the basis of the analysis of the phonation time and the statistical distributions of the values of periods of the larynx excitation and their percentage changes in time. A final evaluation was performed, using a diagnostic model in which fuzzy relations were used for the classification of voices into two sets-normal and pathological voices. The mean classification error did not exceed 8%, with a lack of decision occurring in about 20% of the casses.

1. Introduction

For a number of years an increasingly widespread tendency to use more or less automated measurement systems, permitting an objective evaluation of the state of specific organs of the human organism, could have been observed in medicine. Development of suitable low-cost methods and diagnostic apparatus, permitting mass screenings to be conducted, mainly for preventive purposes and the early detection of disease, is a particularly important problem. In the literature on the investigation of human speech organ there are relatively few papers concerning acoustical methods of voice evaluation, although, compared

---

* The paper was written under the problem 10.4.3 — Biocybernetic systems and elements.

with many traditional methods used in phoniatry, they show many advantages. First of all, they permit the taking of measurements fully or partly automatically, using microprocessors or minicomputers. In addition, they do not require the introduction of devices or sensors into a speech organ or their application outside the organ. When traditional methods are used, e.g. endoscopic, stroboscopic, bioelectrical, radiological etc., the voice examination is carried out under abnormal articulation conditions, which causes considerable uncontrolled variation in the measurements. On the account of the great discomfort to the patient in this type of examination, they are limited to voice evaluation on the basis of an analysis of some vowels uttered in isolation, and do not usually permit an investigation of the action of a speech organ during the utterance of longer texts, which gives more credible and representative results.

The above-mentioned faults of traditional methods for voice analysis are particularly conspicuous in the investigation of the activity of the larynx tone. This investigation is nearly always performed under abnormal articulation conditions and is usually limited to the evaluation of phonation during the articulation of short speech segments. Probably this is one of the reasons why these methods fail to detect early pathological changes occurring in the larynx while they are still small.

The above-mentioned limitations do not apply to the acoustical methods of voice investigation which analyze a complex speech signal measured using a microphone placed at a certain distance from the mouth of the speaker. In addition the investigation of voice using these methods can be performed on the basis of a signal recorded on an audio tape, which facilitates the storage of measurement data and the repetition of analyses as measurement and analytical methods improve.

This paper presents a method for the investigation of larynx phonation based on the analysis of the phonation time, the variation of the frequency of the fundamental tone, and the statistical distributions of the period of the larynx tone and its percentage changes. It also describes a diagnostic model assumed in the investigations for the evaluation of the state of the larynx source, which was verified for specific cases of pathological changes in the vocal chords.

2. Larynx excitation in the case of normal voices

One of the basic anatomical parts of the speech organ, which participate directly in the creation of an acoustical speech signal, is the larynx which is the upper end of the trachea through which the air pressed out of the lungs passes into the mouth cavity. The air stream which flows through the larynx is modulated in the glottal slit formed between the cyclically opening and closing vocal folds. Due to the motion of the vocal folds, air pulses are created, which excite vibrations of the air mass in the throat and mouth cavity. The behaviour of variations of the glottis opening area (and thus the similar variations of the volume velocity of the air stream) is shown schematically in Fig. 1.

Larynx excitation is the dominating kind of excitation of the vocal cavities and is for the most part the primary source of acoustic energy in a speech signal. In addition, the occurrence or lack of a voiced characteristic in physiological speech has phonetical significance, and changes in the frequency of the larynx tone or its amplitude have linquistic significance. Thus in articulation the larynx plays two important functions: (a) a phonetic function which permits sounds to be distinguished as voiced and unvoiced, and (b) a prosodic function, since changes in the intensity of speech and its fundamental frequency serve to stress chosen syllables and to create different melodic contours.

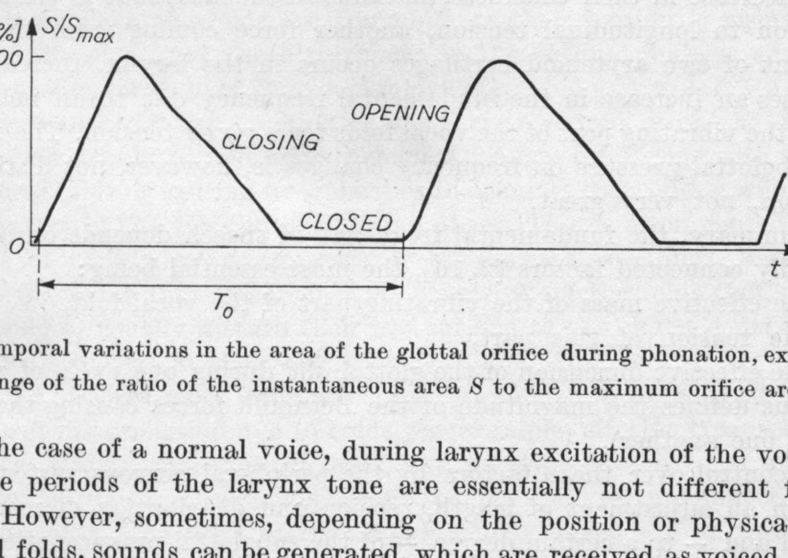


Fig. 1. Temporal variations in the area of the glottal orifice during phonation, expressed by the change of the ratio of the instantaneous area S to the maximum orifice area S_{max}

In the case of a normal voice, during larynx excitation of the vocal tract, successive periods of the larynx tone are essentially not different from one another. However, sometimes, depending on the position or physical state of the vocal folds, sounds can be generated, which are received as voiced, although successive periods of the fundamental tone differ essentially one from another. This can occur in two ways. Most often successive periods are alternately shorter and longer. The periods occurring alternately become increasingly closer to each other (nearly identical) than to the adjacent ones. However, steplike changes occur relatively less often in the duration and shape of the larynx pulses.

As was already mentioned, the larynx source plays a dominant role in the formation of the prosodic features of a speech signal, which are expressed in the behaviour of the melodic contour and in changes of intensity. A third factor affecting prosodic characteristics, duration of the individual speech sounds, which is physiologically conditioned, primarily by the velocity of the supraglottal articulation elements, depends directly on the larynx excitation. The intensity of the output signal is conditioned only partly by the intensity of the larynx source, which depends mainly on the magnitude of the subglottal pressure. If, for example, this pressure increases by a factor of two, the level of larynx excitation increases by about 12 dB [2]. A number of investigations have shown that the most important prosodic characteristic is the behaviour of the melodic contour, while an exact control of changes in the fundamental frequency is essential for the creation of an utterance with a predetermined intonation.

The melodic contour is defined by the variations of the fundamental frequency which is physiologically dependent primarily on the eigenfrequency of the vocal folds. The value of this frequency is affected by such physical parameters of the folds as their length, thickness, mass of vibration, and tension. Thus tension is directed along the vocal ligaments and the frequency of the chord vibrations increases with increasing tension in the ligaments. Also, the lengthening of the folds causes the fundamental frequency to increase, due to a decrease in their thickness, and thus in the mass that is vibrating [8]. In addition to longitudinal tension, another force coming from the central attachment of two arytenoid cartilages occurs in the larynx. Increasing this force causes an increase in the fundamental frequency due to an indirect decrease of the vibrating part of the vocal folds for a given tension. The influence of the subglottal pressure on frequency changes is, however, not distinct and, in addition, not very great.

In summary, the fundamental frequency of speech depends on a number of mutually connected factors [2, 16], the most essential being:

1. the effective mass of the vibrating part of the vocal fold,
2. the tension of this part,
3. the effective dimension of the glottal slit during one cycle of vibration (the radius defines the magnitude of the Bernoulli forces causing the folds to approach one another).

The control over these factors by the peripheral nervous system of man is through an adjustment of length, tension and distribution of mass in the vocal fold and — to a certain degree — of the subglottal pressure affecting the magnitude of the Bernoulli forces.

In pathological cases involving the larynx source, disturbances in their phonation function can occur due to anatomical changes (e.g. the occurrence of polyps) or due to damage in the system which controls its work, e.g. paralysis of the recurrent laryngeal nerves.

3. Disturbances in the function of the larynx source, due to pathological changes in the vocal chords

The investigations presented in this paper involved disturbances of voice caused only by pathological changes in the larynx. Disturbances in its phonation function due to changes in the endocrine glands, in the central nervous system, and personality and emotional changes were not considered.

Most often the faulty operation of the larynx source is a result of a failure of the glottis when periodically closing, to do so in a complete manner, or of asynchronous motion of the vocal folds. The failure of the glottis to close totally and the absence of a rapid break in the flow of the air stream cause the higher harmonics of the larynx tone to have a lower level than in the case where full closing occurs. This is due to the fact that variations in the volume velocity of the stream with incomplete closing of the glottis are considerably smaller than when there is a full closing of the chords, which is necessary for a suitable increase in the subglottal pressure to occur. In addition when, at the moment of the maximum proximity of the folds to each other, the area of the glottal slit is sufficiently large, turbulence can occur in the air flow. The absence of a closing phase for the vocal folds usually causes the voice to be weakened, and often coarse. It is often accompanied by a shortening of phonation time and, in more serious cases when the folds are immobile or their motion is insufficient, by aphonia. Almost always the disturbance in the closing phase is accompanied by asynchronous motion of the vocal folds [11].

Disturbances in the function of the larynx source, from the anatomical point of view, can be caused by pathological changes interfering with the motion of the vocal folds or causing irregularities to occur at their edges, which prevent exact contact of the folds in the closing phase. This is often caused either by paralysis of certain larynx muscles or by growths, swellings, scars, polyps. It is worth noting that larger growths occurring on the folds (not necessarily at their edges) usually disturb their motion due to the changed mass distribution and differences in their elasticity [11], for example.

Of the diseases most often affecting the phonation activity of the larynx all states of inflammation due to colds, for example, effusions (caused by excessive vocal effort), cancerous changes, and paralysis of the vocal folds should be mentioned. Cancerous changes, both benign (polyps, papillomas etc.) and malignant, affect the work of the larynx source, depending on their size and location. When they are not very large and occur on the edges of the folds, the changes in the voice which they cause, can be unnoticed. This is particularly critical in the case of malignant tumours, which can be successfully cured, provided they are detected early enough. Otherwise, it is usually necessary to perform a partial or total laryngoctomy. Adduction and also obduction of the vocal chords can also be limited as a result of paralyses or paresis of the larynx muscles. These paralyses, which are usually unilateral (less often bilateral), are most frequently caused by the damage to larynx nerves during surgical operations on the thyroid. It is a rather complicated operation and even faultless operational technique does not sometimes protect the patient from the damage to larynx nerves which run very close to the thyroid gland, usually in an untypical way.

The work of the larynx source in cases of a unilateral damage depends predominantly on the position of the paralysed vocal fold. If it is close to the central line of the glottis (in the so-called medial position), it can function in an acceptable manner, since the healthy fold reaches the damaged one without difficulty and no changes can be observed in the voice. In the case of a more distant position of the damaged fold, the healthy fold cannot cross the centre line and a full closing cannot thus occur [11].

In general, from the acoustic point of view, malfunction of the larynx source can result in time variations of the fundamental tone, the intensity of the speech signal, the duration of the voiced excitation and the voice quality.

The pitch is determined by the frequency of the larynx pulses which, in the case of pathological changes in the vocal folds, can be subject to considerable random variations, as a result of a decrease in the tension and flexibility of the folds and also of changes in the spatial distribution of their mass. These factors decisively affect the speed of motion of the folds and the glottal closing time. The energy generated in each successive larynx pulse has a dominating effect on the behaviour of the level of speech intensity with time; thus, for example, asynchronous motion of the ligaments or factors affecting the degree of glottal closing play an essential role here, since the excess subglottal pressure generated just before the closing of the folds depends on these. Also, the phonation time, which in the case of incomplete closing of the glottis decreases due to an increase in the rate of flow of air through the larynx, depends on the degree to which the glottis closes. Voice quality, which is defined in a rather subjective manner, depends on the factors mentioned above. In particular, asymmetric motion of the folds is reflected in a hoarseness, weaker sonority and an unnatural sound of the voice, although the closing may be nearly complete.

4. Methods for the investigation of the phonation activity of the larynx source

In laryngology and phoniatry generally accepted clinical methods for voice investigation exist, which can be divided into two basic groups: (a) audio monitoring and (b) visual monitoring. Since structural changes in the larynx are often accompanied by changes in the voice, the performance of suitable articulation tests can supply the physician with essential information [4]. The main fault of audio monitoring methods is their subjective character, which results in more difficult cases when changes in the voice are not distinct, in the possibility of even experienced phoniatrists giving different diagnoses for the same patient. In addition, there is a lack of quantitative standards for audio evaluation of the state of the larynx source.

Visual methods, e.g. endoscopic, stroboscopic, electromyographic, radiological etc., enable the physician to perform a fairly reliable evaluation of the operation of the larynx source in cases with advanced pathological changes. These methods, however, mainly because they are time-consuming and expensive, are not useful for mass examinations which are becoming increasingly necessary for the early detection of disease, particularly those of the cancerous type [5]. In addition, when using these methods it is, in general, impossible to perform a long-term investigation of the phonation activity of the larynx source during the utterance of longer texts, mainly because these must occur under abnormal articulation conditions which are distorted by devices and alien bodies introduced into the speech organ and by the performance of additional tiring activities. These investigations are, therefore, usually limited to the observation of the larynx during the articulation of isolated vowels and, in many cases, when structural changes in the larynx are slight, disturbances in phonation can occur in a random and rather infrequent manner, so that they may pass unnoticed in this shorter observation time.

Acoustical methods, which are free from the faults mentioned above, investigate a speech signal recorded by a microphone placed at a fixed distance from the speaker's mouth. In the case of an investigation of the larynx source, a special measurement methodology must be developed, which permits information to be obtained from the analysis of such a complex signal as the acoustical speech signal. The final form of the signal is influenced not only by the frequency characteristics of the larynx source, but also by transmission characteristics of the throat-mouth (and possibly nasal) tract and by the mouth radiation characteristic.

A basic acoustic parameter that is analysed in the evaluation of the state of the larynx source is the behaviour of the fundamental frequency in the speech. LIEBERMANN [12], who was the pioneer in the investigation of this type behaviour for cases with pathology of the larynx, analysed the distribution function of the differences between successive periods of the larynx tone and introduced a synthetic index – the so-called *pitch perturbation factor* – defining the probability of occurrence of differences in excess of 0.5 ms. Liebermann showed that this index is larger for cases with pathological changes of the larynx than for physiologically normal states. KOIKE [10] and DAVIS [5] used an evaluation method of a similar type, although slightly modified, for the assessment of the state of the larynx source.

The variation in the intensity of the harmonic excitation in a speech signal is another parameter which, it seems, should be more widely used for evaluation of the state of the larynx source. However, the measurement is very difficult and complicated. Davis [5] was one of the first investigators to use an analysis of the intensity variation of the excitation for the evaluation of the phonation activity of the larynx. He investigated the distribution of the amplitudes of the larynx pulses by inverse filtering a signal recorded during a lengthened articulation of an isolated vowel /a/. He found that in the cases with pathology of the larynx, the distribution is usually more diffuse and has a smaller kurtosis than that for normal voices. This is most probably due to the fact that additional larynx noises, caused by turbulent air flow through the glottial slit, often occur in pathological cases.

Recently investigators have attempted to evaluate the pathology of the larynx source by an analysis of the whole spectral structure of the speech signal, and not only chosen parameters connected directly with the source [5, 6, 10]. It appears [5], however, that the analysis of the behaviour of the frequency and amplitude of the larynx excitation gives most information on the phonation activity of the larynx.

Acoustical methods, which are being developed for the evaluation of the voice forming activity of the speech organ, should be assessed from two points of view. They must first of all permit a differentiation between normal and pathological states; and subsequently must provide a quantitative measure of the changes occurring in the voice during the course of the curing and rehabilitation procedures. Further development of these methods will probably be towards the provision of an objective evaluation of the degree of advancement of the pathological states, with particular emphasis on their early detection.

5. Bases of the diagnostic model used in the investigations

The primary task of medical diagnostics is first of all to establish if the person presenting himself for examination is healthy or not. It is only later that the type of disease is identified and a decision can be made about the proper treatment. In the course of therapy, diagnosis serves to follow the advancement of the treatment and permits it to be ended at the proper time.

Automation or, at least, computer-aiding of the diagnostic process, is becoming more and more attractive for physicians and scientists. Most of investigations performed in this field are based, however, on statistical decision models, and the results obtained have not been very promising so far from the point of view of medical applications [19]. The criterion that is most often used in decision-making is Bayes' criterion based on the assumption that the best set of patterns is the one which gives least probability of classification error. In practice, the creation of such a set requires an exact knowledge of the multidimensional distributions of conditional probabilities, which involves the need to possess a sufficiently large data base, the dimensions of which are, more often than not, unattainable.

Thus, it is often necessary to use approximate distributions which usually deviate considerably from the real distributions. In addition, all attempts at expanding the model to include new diseases require the recalculation of all the conditional probability distributions, (based on a wider data base) and a complete change of the decision-making rules. In practice, the precision of classification with different statistical decision-making models is unsatisfactory, and when tested on an open set of data does not usually exceed 70 % [15]. It seems that further development of statistical decision-making models will not bring an essential improvement and it is thus necessary to use a completely different approach to the diagnostic process.

In a classical diagnostic technique used in medicine, the state of the human organism or a particular organ is evaluated by the physician on the basis of a number of symptoms in a manner which is, to a dominating degree, subjective (based on experience acquired in the professional practice). Even those of symptoms that can be measured, e.g. temperature, are described in an approximate manner, using designations: normal, slightly increased, increased, high, very high etc. temperature. Individual symptoms together form a characteristic set and diagnosis consists in selection of the disease most typical for the given set of symptoms of disease. The use here of statistical classification methods necessitates the formation of an exact quantitative description of the disease phenomena investigated, which are naturally described in a more or less approximate way.

In the diagnostic model used in this work it was assumed that the analysis of individual symptoms (features) is expressed in mutually opposing fuzzy sets [20], i.e. less formally in sets, where a sharp change from full assignment to nonassignment to the set does not occur, but where intermediate assignment stages can also exist. In this approach, the degree of assignment (so-called *grade of membership*) is expressed by a pair of numbers in the range [0, 1], the former indicating nonassignment, and the latter assignment to a given set.

It is assumed that N is a subset in the space X. A function $\mu_N(x)$ is defined, called the *characteristic function* of the set N, giving for each element x in the space X the degree of its assignment to the subset N in terms of the number $\mu_N(x)$ in the range [0, 1]. When $\mu_N(x) = 1$, the element x belongs to N, while the condition $\mu_N(x) = 0$ corresponds to the fact that x does not belong to N.

If each element x is determined, using a series of measurable properties $C = C_1, C_2, \ldots, C_n$, it is assigned to a suitable value of $\mu_N(x)$ on the basis of an analysis of variations of the individual parameters, followed by a so-called composite inference rule [20], based in this case on the linguistic characteristics of the individual features.

An example of voice pathology will serve to give a brief explanation of this method of feature description. As was mentioned in Section 2, in most cases of larynx pathology an decrease in the phonation time can be observed with respect to normal voices. If the phonation time is short, the hypothesis that pathological changes occur in the larynx cannot be rejected. If, however, the time is long, it can be assumed that the larynx source functions normally. In the present paper, the ratio of the total duration of voiced excitation to the whole duration of the utterance was used. Theoretically, this ratio can take any value in the range [0, 1], depending on the type of text uttered, the emotional state of the speaker or the type of disease. In the case of aphonia, for example, it is close to zero. In practice, for a normal voice and phonetically balanced spoken text, this ratio does not usually exceed a value of 0.85.

As an example, Fig. 2 shows the behaviour of the assignment function $\mu_L(A)$ for a fuzzy subset of "low values", which takes the following values:

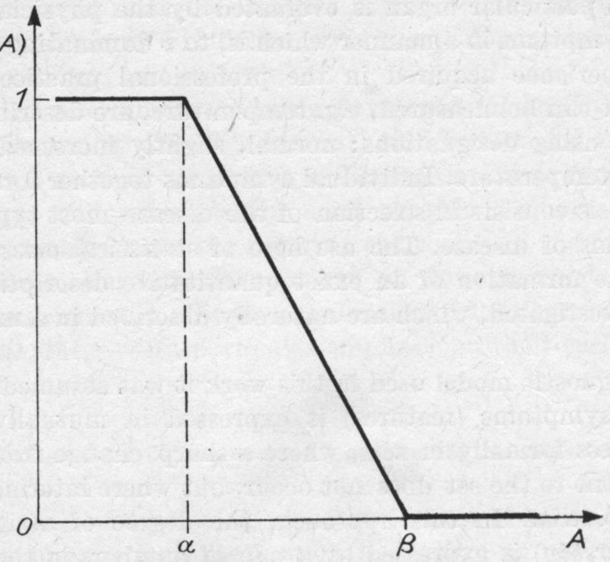


Fig. 2. An example of the behaviour of the assignment function $\mu_L(A)$ of a fuzzy subset "low values" of A

$$\mu_L(A) = \begin{cases} 1 & \text{for } 0 \leqslant A \leqslant \alpha, \\ (\beta - A)/(\beta - \alpha) & \text{for } \alpha < A < \beta, \\ 0 & \text{for } A \geqslant \beta. \end{cases} \tag{1}$$

Thus, in the case of aphonia, $\mu_L(A) = 1$. In an analogous manner, the subset of "high values" can be defined, for which the assignment function $\mu_H(A)$ is connected with $\mu_L(A)$ by the following relation:

$$\mu_H(A) = 1 - \mu_L(A). \tag{2}$$

Determination of the characteristic points α and β of the assignment function $\mu(A)$ is performed as a result of a learning process, based on utterances by normal voices, with long phonation times, and by pathological voices, with short phonation times. The quantity α is defined as that value of the parameter A, below of which the phonation time is low, while that of β is that value above of which its value is certainly high. A similar description is made for the remaining features, with the assignment function of type μ_H or μ_L being determined for each separately. In this manner, the state of the larynx source can be described, using a fuzzy relation of its features, and the degree of assignment of the ith investigated voice to the set of pathological voices or the set of normal voices can be determined, based on the formulae

$$\mu_P(i) = \bigvee_j \mu_{R_p(C_j)}(a_{ij}), \qquad \mu_N(i) = \bigvee_j \mu_{R_n(C_j)}(a_{ij}), \tag{3}$$

where $\mu_P(i)$ and $\mu_N(i)$ define the degree of assignment of the ith voice, to the set P of pathological voices and set N of normal voices, respectively. $R(C_j)$ denotes the fuzzy limit of the value a_j of the parameter A_j, defining a subset of values $\{a_j\}$ having a common property, according to the assumed description (e.g. A_j is *low*); $\bigvee_j$ is the operator of determining the upper limit with respect to j.

The values of the degree of assignment μ_p and μ_N are the criteria for making the decision as to which set an analysed voice will be assigned. The analysed voice is assigned to the set whose degree of assignment is greater; with an additional condition of the form $(\mu \geqslant d) \vee (\mu < 1-d)$, where $d = 0.07$, being introduced before a decision can be made. The final form of the decision-making rule is thus the following:

$$\begin{aligned} P_d &= \{i \mid (\mu_p(i) \geqslant d) \wedge (\mu_N(i) < 1-d)\}, \\ N_d &= \{i \mid (\mu_p(i) < 1-d) \wedge (\mu_N(i) \geqslant d)\}. \end{aligned} \tag{4}$$

When these conditions are not satisfied, the voice analysed was assigned to a "no-decision" set.

6. Measuring system

The frequency of the period of the larynx tone can be measured in many ways and for normal voices (with individual exceptions) the measurement is not usually a very difficult technical problem. In pathological cases, however, the measurement is not easy (see [3]), routine digital methods of larynx tone extraction fail totally and it is necessary to use tiring, semi-automatic procedures. This was one of the reasons why an analogue method, which imposes less limitations on a signal and at the same time permits real time measurements to be made, was used to extract the larynx tone from the speech signal.

In the investigations a measuring system was used which permitted measurement of the following running value of changes in the period (frequency) of the larynx tone, the percentage variation recorded for each period, and the relative duration of voiced excitation (expressed as the ratio of the duration of larynx excitation to the duration of total utterance). A schematic block diagram of the system is shown in Fig. 3.

An Intonograf-77, designed and produced in Acoustics Laboratory of the Institute of Fundamental Technological Research of the Polish Academy of Sciences, whose design and operation has been discussed previously [14], was used to extract and to measure the period of the larynx tone. Two filter systems were used in the device to extract the fundamental tone. One was fixed with a pass band of 78 to 500 Hz, and the other was used as a low pass tracking filter, which retuned automatically so that the fundamental frequency of the speech signal was always slightly above its limiting frequency. The period of a signal thus filtered is determined in a continuous manner by measurement of time between two succesive unidirectional zero crossings of the signal (the measuring precision of a period is 0.05 ms). In the case of the absence of voiced excitation, the signal quantization time automatically assumes a constant value of 12.5 ms, and the register of the device is zeroed.

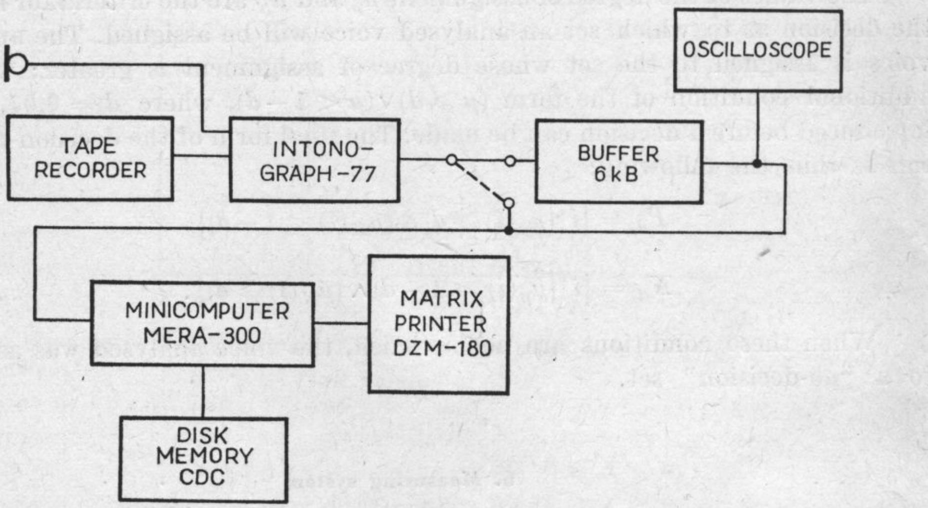


Fig. 3. A schematic block diagram of the measuring system used in the investigation

In the present measuring system, the Intonograf is connected on-line to a Mera-300 minicomputer which has 8-bit words and a memory cycle time of about 1μs. A 7502 Brüel and Kjaer Digital Event Recorder was used in the system for buffering the digital data from the Intonograf. It was thus possible to check the data visually on an oscilloscope screen before sending it to the minicomputer. The software of the measuring system consists of a series of programmes for plotting the frequency variation of the larynx tone with time, for plotting statistical distributions of the values of the periods of larynx excitation and of the values of the relative differences between successive periods, and for the determination of distribution parameters such as the mean values, various dispersion measures, coefficients of asymmetry and kurtosis, modes, medians etc. The relative differences between successive periods of the larynx tone were determined according to the formula

$$\Delta TO = \frac{2(TO_i - TO_{i+1})}{TO_i + TO_{i+1}} \tag{5}$$

for $TO_i \wedge TO_{i+1} \neq 0$, where i is the number of the (successive) periods measured.

The distribution of the values of ΔTO can be considered, as an approximation, to be equal to the distribution of the values of the derivative of the function $TO(t)$ with respect to time.

To date this has not been discussed in the literature.

Part of the programmes was written in assembler, while the remainder (concerning statistical analysis) were written in a limited version of Fortran adapted for the minicomputer. The plot of the variation of F_0 was made immediately at the end of a measuring cycle, and the statistical processing of the data lasted about 5 min. An example of the printout of the results of the statistical calculations and the distributions of TO and ΔTO [%], obtained for a voice with polyp changes of a vocal fold, is shown in Fig. 4.

7. Investigation material

The investigations were performed on 143 persons, including 54 with normal voices (26 men and 28 women), 54 persons (including 37 women) with paralysis of the larynx, caused by injury to the recurrent nerves following operation on the thyroid gland, and 18 persons (including 13 women) with polyp changes on the vocal folds. The other disease states included chronic laryngitis, angiomas, larynx cancers, etc. In addition, for 12 of the subjects, several measurements were made at different stages of the treatment process and in the course of rehabilitation. Prior exact laryngological and phoniatric examinations were performed on all patients to exclude all other causes (outside the larynx) which could essentially affect the respiratory and voice-making functions of the larynx.

Healthy subjects also underwent exact phoniatric examinations.

The material used for measurement and analysis was a speech signal recorded on audio tape, of the individual utterance of two texts, successively: (a) a text of three pairs of sentences with the same phonetic structure and differing only in the intonation pattern (statement — question), and (b) a so-called newspaper text, i.e. part of a press article of adequate length to be phonetically balanced. The reading time of the second text lasted about 1 minute.

8. Analysis of the measurement results and the selection of parameters for a description of the condition of the larynx source

On the basis of the results obtained previously [7], the present investigations concentrated on the analysis of the intonation contours in questions and on the statistical distributions of TO and ΔTO, determined for a newspaper text.

```
LEOKADIA S          L.51
POLIPOWATY TWOR PRZY LEWYM FALDZIE NALEWKOWYM
NAGR. 30.05.1975

                    CZAS (MS)    (%)    PROBKI   (%)
    CALA PROBA       44862.54    100      6144   100
    BRAK SYGNALU     22124.99    49.3     1770   28.8
    ODRZUCONE           50.89     0.1       21    0.3
    SYGNAL           22686.64    50.5     4353   70.8
    ODRZUCONO PROBKI DLA T <   1.69 MS   ORAZ DLA T > 12.69 MS

    PARAMETRY ROZKŁADU TØ

                       SREDNIA =   5.21 (MS)   (191.87 HZ)
                     WARIANCJA =   2.25
         ODCHYLENIE STANDARDOWE =   1.50 (MS)
                     NORMOWANE =   0.2878

    PARAMETRY CENTRALNE        ASYMETRIA =   2.1401
                                  EKSCES =   6.8302

    PARAMETRY POZYCYJNE        MODA 1 :    4.52 %  DLA    4.89 MS
                               MODA 3 :    4.38 %  DLA    4.89 MS
                               MODA 5 :    4.20 %  DLA    4.94 MS
                    MEDIANA :   4.94 MS   (202.02 HZ)
        KWANTYLE       P.10 :   4.04 MS
                       P.25 :   4.64 MS
                       P.75 :   5.34 MS
                       P.90 :   6.29 MS
                        ROZPROSZENIE =    0.34 MS
                           NORMOWANE =    0.0707
               ASYMETRIE   P.10-P.90 =    0.1999
                           P.25-P.75 =    0.1428
                              EKSCES =    0.1555

    PARAMETRY ROZKŁADU DTØ

    PRZETWARZANO WARTOSCI DTØ W ZAKRESIE OD  -150   DO   150  %
                       SREDNIA =   0.13 (%)
                     WARIANCJA = 668.42
         ODCHYLENIE STANDARDOWE =  25.85 (%)

    PARAMETRY CENTRALNE        ASYMETRIA =  -0.2591
                                  EKSCES =   5.3507

    PARAMETRY POZYCYJNE        MODA 1 :    4.43 %  DLA DTØ =    0
                               MODA 3 :    4.31 %  DLA DTØ =    1
                               MODA 5 :    4.18 %  DLA DTØ =    1
                    MEDIANA :    0.00  %
        KWANTYLE       P.10 : -22.99
                       P.25 :  -6.99
                       P.75 :   7.99
                       P.90 :  22.99
                        ROZPROSZENIE =    7.50
               ASYMETRIE   P.10-P.90 =    0.0000
                           P.25-P.75 =    0.0666
                              EKSCES =    0.1630
```

Fig. 4 a

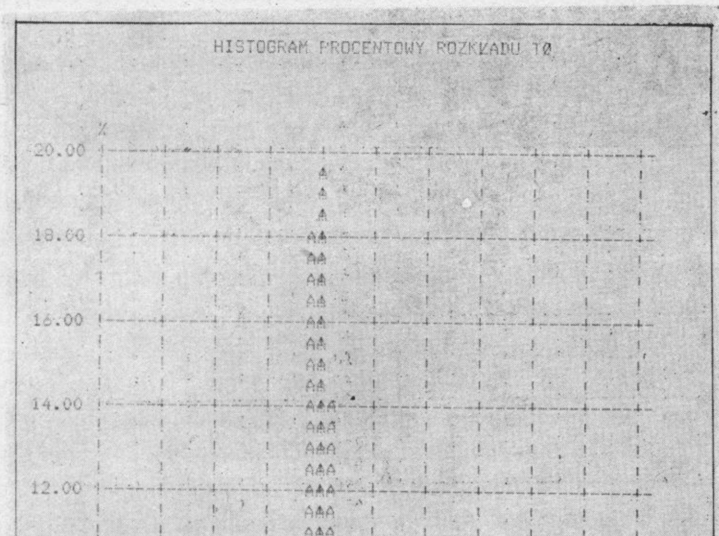

HISTOGRAM PROCENTOWY ROZKŁADU TØ
%
20.00
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
1.25 2.5 3.75 5 6.25 7.5 8.75 10 11.25 12.5 (MS)
800 400 250 200 160 125 100 80 (HZ)

Fig. 4b

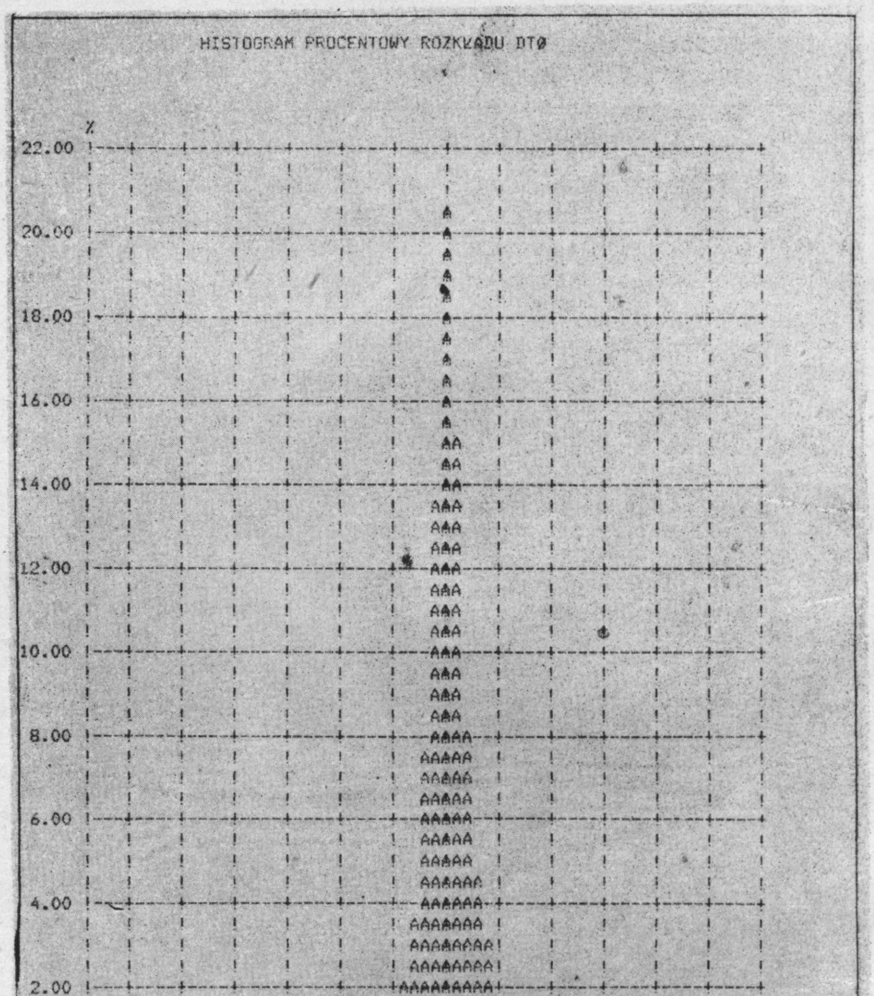


Fig. 4c

Fig. 4a-c. A printout of the results of the statistical calculations with plots of the distributions of *TO* and *TO* obtained for a voice with polyp changes on the left-hand vocal fold

For a general evaluation of the state of the larynx source, both at the time of reporting and in the course of the patient's rehabilitation, an analysis of the melodic contour of recorded questions can give the phoniatrist much information. Firstly, it can easily be estabilished the regularity ("smoothness") of tem of the measured larynx tone. Secondly, it can readily be observed the degree to which the patient is able to raise the intonation at the end of a sentence. In the first case, the occurrence of relatively large steplike changes in the frequency F_0 can indicate asynchronous vibration of the vocal folds (e.g. due to paresis, polyps ets.). In the second case, the inability to raise the intonation at the end of a question results from an unnaturally increased tension in the muscles of the vocal folds, which, as a rule, occurs in all the more serious pathological changes of the larynx.

As an example, Fig. 5 shows the variation of the frequency of the larynx tone in a word "domu" uttered in the question: "Czy idziemy do domu?" /t͡ʃɨ id͡zemɨ do domu/ measured on a female voice with a paralysed right vocal fold in the near-centre position. This variation was recorded during the patient's first visit (65a) and one month later, in the course of rehabilitation (65b). In both tests the patient was able to raise the intonation at the end of the phrase, with the increase being more distinct in the second test (about 50% as opposed to 30%). However, in the case considered it is more significant that immediately after the initial increase in the frequency F_0, a smooth fall in the final part of the phrase, together with numerous, rather large steplike changes (of the order of 10-15 %) between successive periods of the larynx tone occurred. Thus, the poor condition of the larynx source was only evident at the end of phonation.

Analysis of this type of behaviour is very convenient, since it gives a phoniatrist a general image of the state of the larynx source, viewed from the point of the acoustical signal generated. Current visualization of frequency changes in the larynx tone can be a valuable aid for the rehabilitation of the voice, for learning the correct formation of phrases with suitable melodic contours with preset intonation. However, when an unambiguous quantitative evaluation of the state of the larynx is required, this type of behaviour is less convenient and requires tedious analyses, which excludes the possibility of using it for mass examination.

For this reason, the analysis of the statistical distributions of selected physical quantities, connected in an unambiguous manner with the character of larynx excitation, is more suitable for the quantitative evaluation of the state of the larynx source. These quantities may be, for example, the fundamental frequency connected directly with the vibration frequency of the vocal chords and the amplitude of the larynx pulses, which, however, is very difficult to determine unambiguously for a continuous speech signal. Such a procedure is easier in the case of vowels spoken in isolation (cf. [6]). However, the deter-

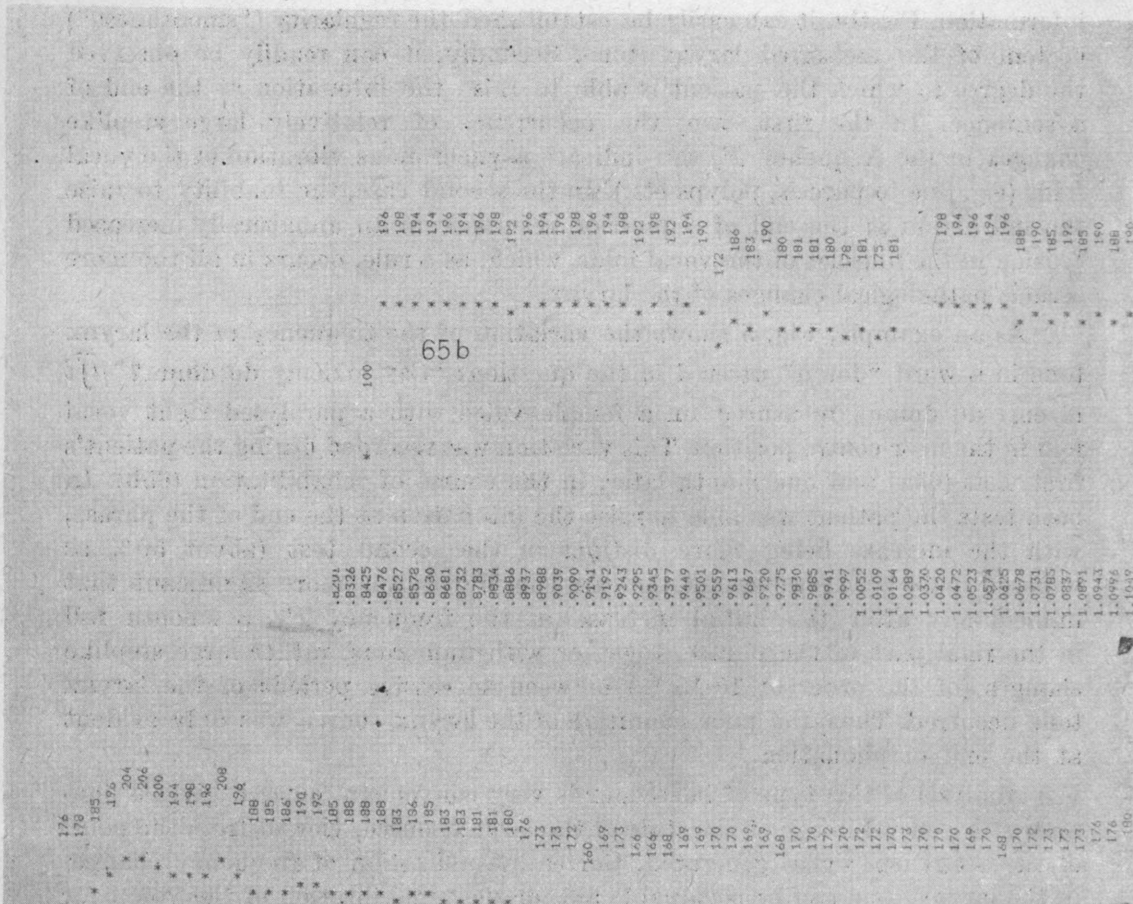
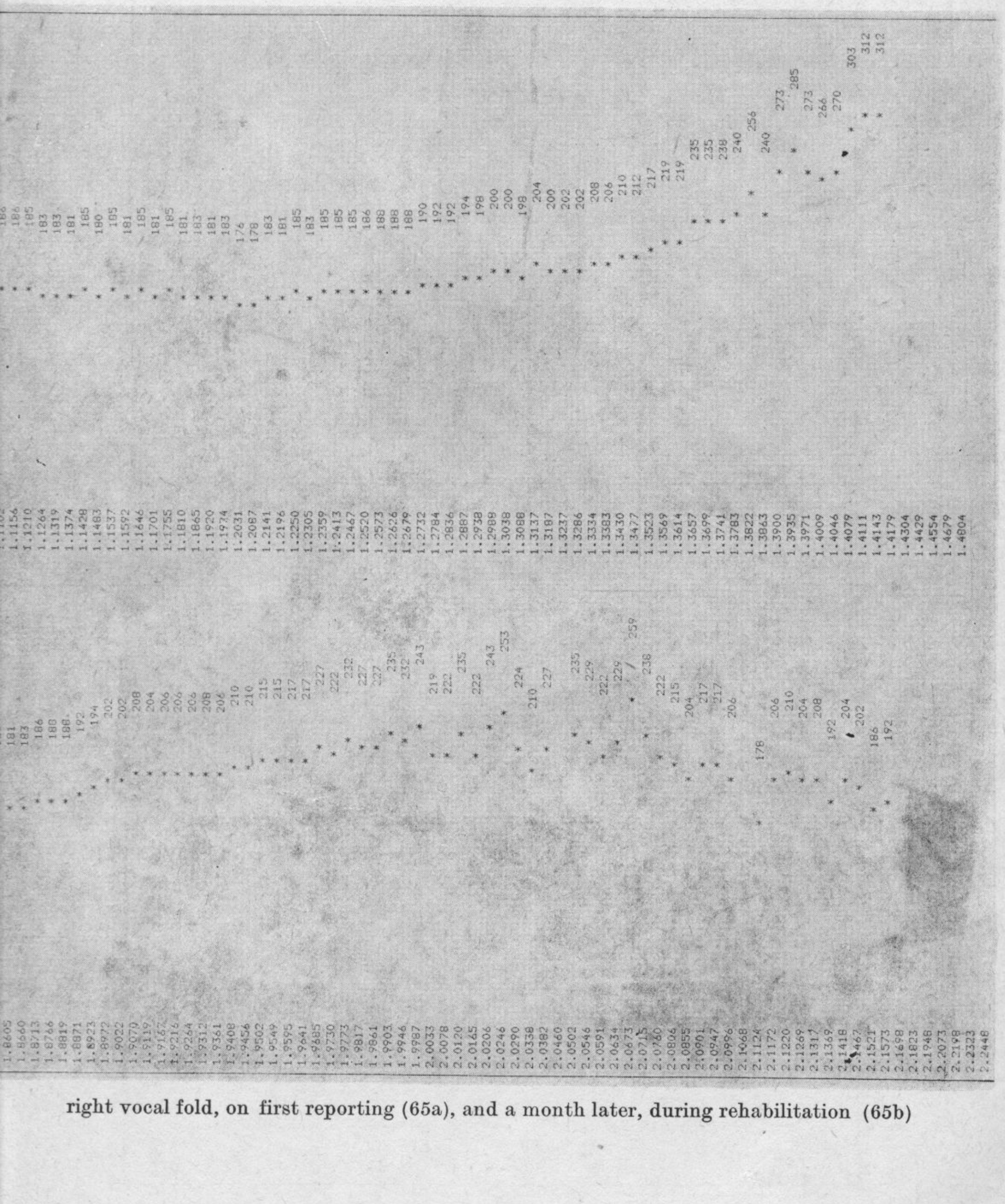
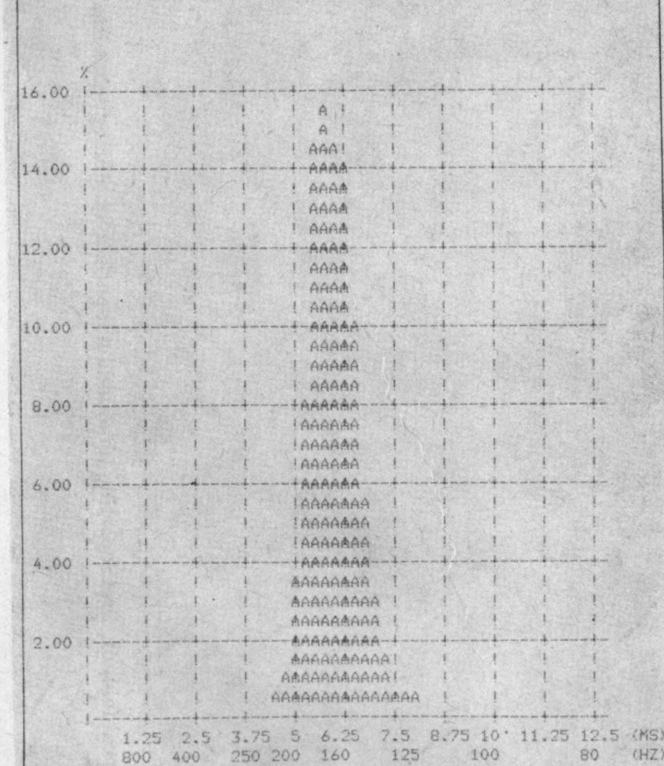


Fig. 5. Temporal variations of F_0 in the word /domu/ uttered by a female voice with a paralysed

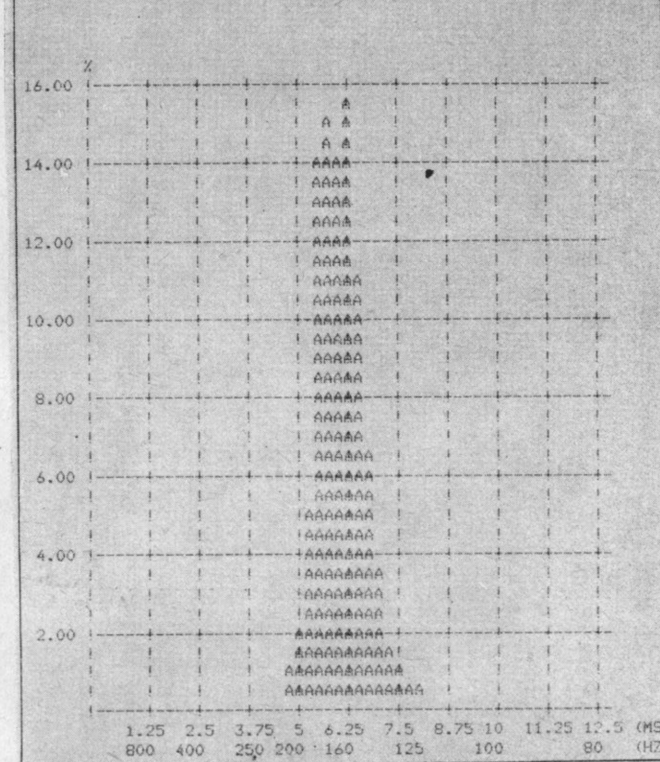


right vocal fold, on first reporting (65a), and a month later, during rehabilitation (65b)

41a


HISTOGRAM PROCENTOWY ROZKŁADU TØ
%
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
1.25 2.5 3.75 5 6.25 7.5 8.75 10 11.25 12.5 (MS)
800 400 250 200 160 125 100 80 (HZ)

41b


HISTOGRAM PROCENTOWY ROZKŁADU TØ
%
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
1.25 2.5 3.75 5 6.25 7.5 8.75 10 11.25 12.5 (MS)
800 400 250 200 160 125 100 80 (HZ)

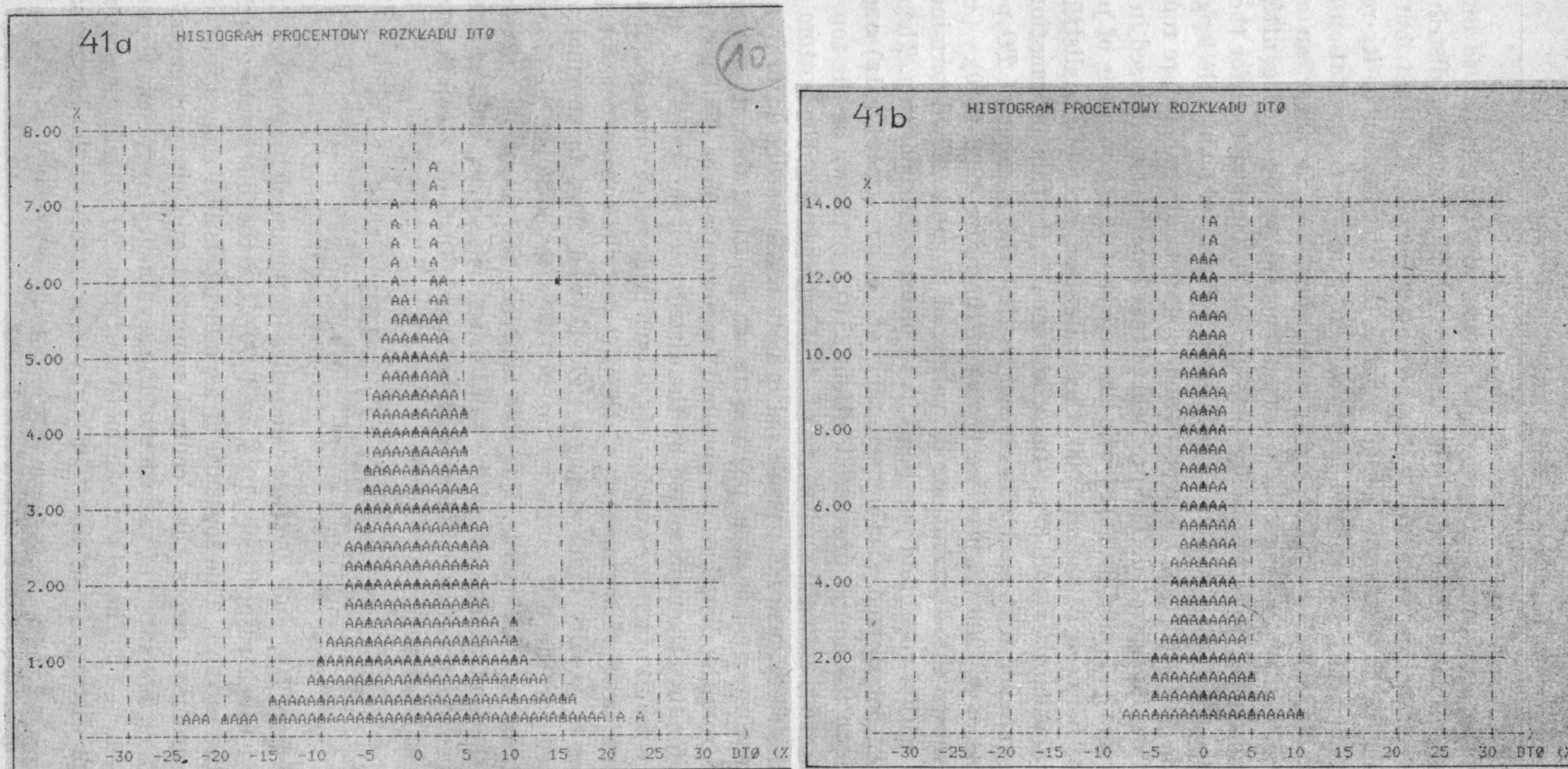


Fig. 6. The distributions of the values of *TO* and *ΔTO* for a male voice with a paralysed left vocal fold, determined at the beginning of treatment (41a) and five months later (41b)

mination, of the frequency (or period) of the larynx tone is not so difficult.

This paper is thus limited to the analysis of the statistical distributions of only one parameter connected with the phonation, i.e. the distribution of the values of the fundamental period T_0 and the distribution of its relative changes in time, ΔT_0, determined for each period of voiced excitation according to formula (5). The second distribution describes, to a certain extent, the stationarity of the changes in the fundamental period during utterance. It is essential during statistical analysis to choose the proper text for the utterance recorded, so that it is phonetically balanced and of a suitable length for it to be safely assumed that the results of the statistical analysis are representative for voices investigated. The problem of text length was analysed in detail by RAMISHVILLI [17] and BOË [1] who showed that if the number of periods measured lies between 2400 and 5000, the evaluation error of the distribution parameters does not exceed 3%. In accordance with this information, the length of newspaper text read was chosen so that the number of *TO* periods measured was over 4000 (in practice it was of the order 4200-4600).

Fig. 6 shows the distributions of values of T_0 and ΔT_0 determined for a male voice with a vocal fold paralysed in an intermediate position. The distributions of T_0 determined at the beginning of the treatment (41a) and five months later, in the course of rehabilitation (41b), are shown at the top of the figure. The corresponding distributions of ΔTO are shown at the bottom of the figure. The distributions of T_0, presented in both cases, hardly differ from each other and have nearly the same parameters. In addition, they do not deviate significantly from the distributions obtained for normal voices. This last remark is also valid for the case represented by the distribution shown in Fig. 4.

Only as a result of a more exact analysis on a large set of voices did it appear that the coefficient of asymmetry of the *TO* distribution (which for pathological voices is usually more strongly right-skewed than for normal voices) could be used for the differentiation of voices with pathological changes in the larynx from normal voices. For this analysis, the traditional measure of the ratio of the central moment of the third order to a square root of the third power of the second order moment was not used. Instead, a coefficient defined [18] by the formula

$$As = \frac{D_{0.9} + D_{0.1} - 2\mathrm{ME}}{D_{0.9} - D_{0.1}}, \tag{6}$$

was used, where $D_{0.1}$ and $D_{0.9}$ are quantiles determining points on the *TO* axis below of which respectively 10% and 90% of the distribution is found, and ME is the median. This measure of asymmetry is less susceptible to incidental measurement errors than central moments and, in particular, those of higher orders. The range of variation of the asymmetry coefficient thus determined is $[-1, +1]$. The other parameters of the *TO* distribution appeared less useful for distinguishing voices in terms of the state of the larynx source.

However, the distribution of ΔTO gives a great deal of essential information on the phonation function of the larynx. A relatively large dispersion of the distribution suggests unstable behaviour in $T_0(t)$ and that numerous, steplike changes of large amplitude occur in it. Also, the asymmetry of this distribution can be used for differentiation of voices. In the example in Fig. 6 it can be observed that the dispersion of the ΔTO distribution after five months of rehabilitation has distinctly decreased (by nearly 50 %). The standard deviation or the coefficient defined from the following formula [18] can be used as a measure of the dispersion:

$$R = \frac{D_{0.75} - D_{0.25}}{2}, \tag{7}$$

where $D_{0.25}$ and $D_{0.75}$ are the quartiles of distribution, defined in an analogous way to the quantiles occurring in formula (6). The investigations showed that the R coefficient is slightly better for the differentiation of voices than the standard deviation. This coefficient does not usually exceed a value of 0.05 for normal voices. In the case considered the asymmetry also decreased a little, although it is less visible on the plots shown.

Three parameters connected with distributions analysed, i.e., the asymmetry coefficients for the TO and ΔTO distributions and the coefficient R of dispersion in the distribution ΔTO, were finally used for the description of the state of the larynx source. In addition, another parameter, the relative phonation time τ_{ph}, which has been discussed more fully in Section 5, was used for the evaluation.

9. Classification of voices according to the diagnostic model assumed

The model decribed in Section 5 was assumed for the classification of voices into normal and pathological, in terms of the state of the larynx. Each voice was defined, using the four parameters described in Section 8. For each of them separately, the assignment function was heuristically determined for the set of low values (see Fig. 2), at the same time assuming that for normal voices the dispersion R is small, the relative phonation duration τ_{ph} is long, and that the right-skewed asymmetry coefficients of both distributions are small. The values of the parameters corresponding to points α and β were determined for each assignment function, based on the analysis of data from normal voices and voices with paralysis of the larynx. It was found necessary to determine these functions for male and female voices separately. These values are given in Table 1.

On the basis of the behaviour determined for the assignment function, the corresponding degrees of assignment of a given voice to a set of voices with pathological changes in the larynx were determined from the values of the individual parameters. Thus each utterance was described using four degrees of assignment.

The process of classification had two stages. In the first stage only two parameters, R and τ_{ph}, were considered, with the decision-making rule defined by formula (4) being used. When the conditions of the assignment to one of the sets were not satisfied, the two remaining parameters were added to the

Table 1. Values of α and β for the assignment functions of the individual parameters

voices		parameter			
		R	τ_{ph}	As_{TO}	$As_{\Delta TO}$
male	α	0.01	0.5	−0.05	−0.6
	β	0.07	0.7	0.3	0.6
female	α	0.03	0.5	0	−0.8
	β	0.1	0.7	0.4	0

decision-making rule. The classification precision thus obtained was 72 % while a lack of a classification decision for the assignment of a given voice to one of the sets occurred in about 20 % of cases. Detailed results of the classification of the voices are shown in Table 2.

Table 2. Results of the classification of voices from the point of view of the condition of larynx

voices	standard			paralyses			total pathology		
	n.d.	i.	c.	n.d.	i.	c.	n.d.	i.	i.
male	19 %	–	81 %	15 %	15 %	70 %	16 %	19 %	65 %
female	21 %	–	79 %	26 %	5 %	69 %	24 %	7 %	69 %

n.d. – no decision, i – incorrect classification, c. – correct classification

It can be seen from the results shown in Table 2 that better classification results were obtained for female voices. This is particularly visible for the pathological voices and is caused by the limitations imposed by the Intonograf whose lower measuring frequency is 78 Hz, while in many cases the instantaneous value of the frequency of the larynx tone of males reaches as low as 40 Hz and this does not only occur for pathological voices (see [9]). For this reason the investigations did not include very low voices and voices which had a distinct "cut-off" on the right-hand side of the TO distribution. The lower limitation of the measuring frequency in these cases caused a decrease in the dispersion in both distributions and a decrease in the right-skew of the TO distribution. Nevertheless, in some cases errors of this type could not be avoided and, probably for this reason, the pathological voices were classified as normal voices.

As was already mentioned, assignment functions were determined heuristically, based on the results of measurements taken of the parameters of the distributions and τ_{ph}, obtained for normal voices and voices with paralysis of the larynx. The same functions were used for the evaluation of the state of the larynx source for other kinds of pathology of the vocal chords, e.g. polyps, papillomas, cancers etc. From the data in Table 2 it can be seen that the expansion to include these other cases of pathological changes in the vocal folds caused a slight increase in the classification errors. This occurred mainly in the cases of polyp changes for which the phonation function of the larynx depends decisively on the size of the polyp and its position on the vocal chord. If it occurs on the outer side of a vocal fold and is not too large, in practice no changes in the voice and in the behaviour of the larynx tone can be observed.

The present method for the description of the condition of the larynx source is also very convenient for the objective evaluation of changes in the voice occurring in the course of treatment and rehabilitation. As an example, Table 3 shows the results of measurements on a female voice with a paralysed

Table 3. Measurement results for the chosen parameters of the distributions of TO and ΔTO and the relative time τ_{ph}, at different stages in the treatment and rehabilitation of a female voice with unilateral paralysis of the vocal fold

recording No.	parameter				
	$\bar{F}_0$ [Hz]	R [%]	τ_{ph}	As_{TO}	$As_{\Delta TO}$
I	184	4.5	0.46	0.3	−0.04
II	236	10.5	0.64	0.3	−0.09
III	244	7.5	0.65	0.44	−0.15
IV	240	7.5	0.69	0.29	−0.2

right vocal fold in an intermediate position. Measurements were taken when the patients first reported (I), after an operation consisting of a "Teflon" injection into the paralysed vocal chord (II), and twice during the rehabilitation of the voice: three months after the operation (III) and a year later (IV). Table 4 shows the values of the degrees of assignment of the voice to the group of pathological voices, calculated for the individual parameters according to formula (1) for the values of coefficients α and β in Table 1.

It can be seen from the results given in Tables 3 and 4 how the phonation function of the patient's larynx gradually improved in the course of treatment and rehabilitation, although it never reached a perfect condition and the process of voice rehabilitation had to be continued. A distinct improvement occurred first of all in the control of the air flow through the larynx during phonation, particularly after the "Teflon" injection into the paralysed vocal fold, due to which a full closing was attained during larynx excitation. Conclusions concerning changes of this type can be drawn on the basis of an increase in the value of the parameter τ_{ph} obtained on the basis of an increase in the value of the parameter τ_{ph} (and, at the same time, a decrease in the degree of the assignment to the pathological voices). However, the other parameters did not improve as much after this operation, and some even became worse. Audio evaluation

Table 4. Values of the degrees of assignment to a set of pathological voices for a female voice with unilateral paralysis of the vocal fold, determined at different stages of treatment and rehabilitation, based on the data in Table 3

recording No.	parameter			
	τ_{ph}	R	As_{TO}	$As_{\Delta TO}$
I	1	0.22	0.75	0.95
II	0.3	1	0.75	0.89
III	0.25	0.64	1	0.81
IV	0.05	0.64	0.72	0.75

confirmed bad phonation of the patient who after the operation (II) spoke with more effort, and in subsequent trials the voice was a little hoarse and only a slight improvement was observed. It is worth noting that, as a rule, after operations on the vocal folds (e.g. to remove polyps) a worsening of the phonation can be observed in the first stage (speaking with greater effort, hoarseness etc.), which improves only as a result of a very long voice rehabilitation. However, improvement in the closing of the folds occurs very soon after the operation and this is reflected in an increase of the relative phonation duration τ_{ph}.

10. Conclusions

The method suggested for the analysis of the distributions of values of the periods of the larynx tone TO, its relative percentage changes ΔTO, and the relative phonation time τ_{ph}, can be used for initial, objective evaluation of the state of the larynx source. In addition, it can also be used for the observation of changes occurring in it during the processes of treatment and rehabilitation. The introduction of fuzzy relations into the description of the state of the larynx source permits easy assessment of its function.

It is an advantage of the present analytical method that a possible change in the decision-making rules in the case of expansion to include other pathological entities is very easy and does not require the laborious calculations.

The method presented is fully automatic and useful for mass examinations based on the recordings of voices on a tape recorder.

The precision of the classification of voices into normal and pathological was about 72 %, with the mean classification error not exceeding 8 %. It is worth noting here that all healthy voices were properly classified.

The measuring system presented can serve to assist phoniatric diagnosis, and visualisation of the variations of the larynx tone in time during the utterance of sentences with a specific type of intonation can be used not only when a diagnosis is made, but also during the process of the rehabilitation of the voice and speech. The patient can then directly observe the intonation of his utterance and compare it with the given one, chosen for exercises.

Acknowledgement. The recordings of voices used in the investigation were made in the Otolaryngological Clinic Phoniatric Laboratory of the Medical Academy, Warsaw, headed by doc. dr. hab. med. W. Tłuchowski. The authors wish to express their gratitude for the possibility of their use together with medical documentation.

References

[1] L. J. Boë, M. Contini, H. Rakotofiringa, Etude statistique de la fréquence laryngiènne, Phonetica, **32**, 1-23 (1975)

[2] D. J. Broad, *Phonation*, in *Normal aspects of speech hearing and language*, ed. F. D. Minifie, T. J. Hixon, F. Wiliams, Prentice-Hall, Englewood Cliffs 1973.

[3] M. J. Cheng, L. R. Rabiner, A. E. Rosenberg, C. A. McGonegal, *Some comparisons among several pitch detection algorithms*, Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Philadelphia Pa. 1976, 332-334.

[4] F. L. Darley, D. C. Spriestersbach, *Diagnostic methods in speech pathology*, 2nd ed., Harper and Row Publ., New York 1978.

[5] S. B. Davis, *Computer evaluation of laryngeal pathology based on inverse filtering of speech*, SCRL Monograph 13, Speech Communications Research Laboratory, Inc., Santa Barbara, 1976.

[6] B. Frøkjaer-Jensen, S. Prytz, *Registration of voice quality*, Technical Review Brüel & Kjaer, **3**, 3-17 (1976)

[7] R. Gubrynowicz, W. Mikiel, P. Żarnecki, *Acoustical analysis for evaluation of laryngeal dysfunction in the case of vocal cords paralysis*, in *Speech analysis and synthesis*, ed. W. Jassem, vol. 5 (in press).

[8] K. Ishizaka, J. L. Flanagan, *Synthesis of voiced sounds from two-mass model of the vocal cords*, Bell System Techn. J., **51**, 1233-1268 (1972).

[9] W. Jassem, M. Steffen-Batóg, S. Czajka, *Statistical characteristics of short-term average F_0 distributions as personal voice features*, in *Speech analysis and synthesis*, ed. W. Jassem, vol. 3, PWN, Warsaw 1973, 209-225.

[10] Y. Koike, *Application of some acoustic measures for the evaluation of laryngeal dysfunction*, Studia Phonologica, **7**, 17-23 (1973).

[11] H. von Leden, P. Moore, R. Timcke, *Laryngeal vibrations. Measurements of the glottic wave.* Part III; *The pathologic larynx*, A.M.A. Archives of Otolaryngology, **71**, 1, 16-35 (1960).

[12] P. LIEBERMAN, *Some acoustic measures of the fundamental periodicity of normal and pathologic larynges*, JASA, **35**, 244-353 (1963).

[13] W. MIKIEL, R. GUBRYNOWICZ, W. HAGMAJER, *Intonograf – system for measurement and visualisation of changes in intensity and melody of speech* (in Polish), Proc. XXIV Open Sem. on Acoustics, vol. I, Gdańsk–Władysławowo 1977, 120-123.

[14] A. NIEDŹWIECKI, W. MIKIEL, *Digital measurement and coding of fundamental frequency in a harmonic speech signal*, in *Speech analysis and synthesis*, ed. W. JASSEM, vol. 4, PWN 1976, 293-300.

[15] E. PATRICK, *Review of pattern recognition in medical diagnosis and consulting relative to a new system model*, IEEE Trans. on System, Man, Cybernetics, **SMC-4**, 1 (1974).

[16] W. H. PERKINS, *Speech pathology. An applied behavioural science*, Mosby, Saint Louis 1971.

[17] G. S. RAMISHVILI, *Rechevoy signal i individual'nost' golosa*, Mecniereba, Tbilisi 1976.

[18] L. SACHS, *Angewandte Statistik; Planung und Auswertung, Methoden und Modelle*, Springer Verlag, Berlin 1974.

[19] H. WECHSLER, *A fuzzy approach to medical diagnosis*, Int. J. Bio-Medical Computing, **7**, 191-203 (1976).

[20] L. A. ZADEH, *A fuzzy-algorithmic approach to the definition of complex or imprecise concepts*, Int. J. Man-Machine Studies, **8**, 249-291 (1976).

Received on May 21, 1979