SOME REMARKS ON ACOUSTIC PARAMETERS IN SPEECH DISORDERS

Nils H. Buch[1]
Børge Frøkjær-Jensen

1. Introduction

This paper and the following one deal with automatic sampling of acoustic data derived from speech. The actual project is in a preliminary stage.

The first paper presents some reflections about the extraction of the acoustic parameters and the applications to different voice disorders. The next paper describes the data collecting system for the investigation.

1.1. Diagnostics based upon recordable criteria?

Within the fields of hearing pathology the diagnostics has for several years been based on so-called objective criteria - first of all on audiometry. Otologists have in this way been fortunate in possessing methods that could be used for testing the patients, e.g. by means of the patients' responses to given stimuli. Audiometry has been used in its present form for more than 20 years and is now a well-known and well-established method for routine testing of hearing.

In the last few years this has been supplemented with a great deal of investigations on the distinct characteristics of different functions of the ear. Measurements of the impedance in the tympanic membrane, the airpressure, tone-decay, recruitment, etc., and especially the development within the

---

1) Dr. med. Nils H. Buch is in charge of the Phoniatrics Laboratory, ENT Department, of the University Hospital, Copenhagen, and lectures on speech disorders at the Institute of Phonetics.

field of the ERA-audiometry[2] has placed audiology at a central position in modern research.

Within the field of speech pathology it is much more difficult to set up recordable criteria for the defects. The phoniatrics and logopedics still base their diagnosis mainly on subjective estimates combined with laryngoscopic examination by means of the larynx mirror - in recent years to some extent combined with a stroboscopic light source. This poses few problems for the phoniatric doctors who have trained their ears for years, but the method is very uncertain for doctors or logopedics without such experience and without knowledge of the auditive impression of the different speech disorders.

In fact, typical changes in the acoustic speech spectra are often the sole symptoms of many different voice or laryngeal diseases - at least in the initial stage.

These acoustic changes of the speech spectrum have given rise to a very confusing terminology, and such denotions as weak, breathy, noisy, harsh, husky, hoarse, shrill, dull, etc. have been used in the literature for different voice qualities, in most cases without any definition of the words. Furthermore, these characteristics are mostly based upon subjective descriptions of the human voice qualities. The terms do not refer to well-defined standards such as synthetic vowel spectra, which could be very useful in this case. However, in the modern phoniatric examination etiological diagnosis is used in combination with the acoustic/auditive description.

It seems that an international terminology is needed (Sonninen and Damsté 1971, Wendahl 1963, 1966, Isshiki 1966), and that this terminology must be based on recordable characteristics in the acoustic spectrum. This demands a basic

2) The ERA-audiometry is a recordable and objective registration consisting of a summation of the cerebral electric potentials which occur after repeated acoustic stimulation of the ears.

research connecting the different morbi acoustically. It is notorious that this could be done, as proved by the fact that the experienced doctor or speech therapist is able to base much of his diagnosis upon the auditory impression. Moreover, a diagnosis based on analyses of the speech spectrum would be easy to carry out for the nurse and would not be uncomfortable for the patients.

1.2. Phoniatric and logopedic research based on phonetics

The methods and instrumentation of modern experimental phonetics in many respects provide a good frame for applied phoniatric and logopedic research. The phonetician deals with basic research on normal and healthy voices, and he is therefore able to supply the research on voice disorders with his experience and with the necessary reference material for pathologic research. In order to get a sufficiently good reference frame, a great deal of acoustic data of normal speech are needed before it is possible to deal with the acoustics of dysfunctions. It seems to be one of the tasks for the phoneticians to provide doctors and speech therapists with this normative material.

It is obvious that the methods and instrumentation which are used in acoustic analyses of normal voices could be very useful in the analysis of the pathological voice quality as well. During the last years such analyses have been made (Smith 1961, Wendahl 1963, Lauritzen and Frøkjær-Jensen 1970). However, as the measurements normally have been made by hand from curves such as sonagrams and mingograms, the analyses have been of very little use in the clinical situation.

The first demand on a phonetic/acoustic analysis used for the clinical diagnostics must be that the analysis can be carried out automatically as a routine investigation.

The second demand must be that the results must be available shortly after the microphone recordings have been made. Furthermore, it would be an advantage if the results are presented in a form which is easy to handle in the total clinical estimate.

1.3. Collection of data for the comparative analysis

As mentioned above it is necessary to collect a normal material which is adequate for the comparisons between pathologic and normal voices. On this basis it will be possible to compare the acoustic parameters collected from the different voice disorders with those of normal voices.

With reference to the above-mentioned paper (Lauritzen and Frøkjær-Jensen 1970) it is furthermore of practical interest for the value of logopedic therapy to compare the results before and after voice training.

2. The change in voice quality for some typical disorders

The planning of this project has been based upon the following specific speech disorders: vocal mutation, laryngeal paralysis, voice disturbances after androgen therapy (treatment of women with masculine hormons), vocal nodules, cronical laryngitis, and psychogenic dysphonia. The treatment of these disorders can be either pedagogical, medical or operative.

2.1. Vocal mutation

In this physiologically conditioned type of dysphonia the intonation range is normally reduced, and shifts between the voice registers happen very often. The phonation is in many cases pneumophonic.

2.2. Laryngeal paralysis

The voice of patients with a paralysis, most commonly caused by damage of the recurrens nerve, sounds airfilled, noisy, soft, weak, and hoarse. Occasionally the voice is unstable with a tendency to diplophonia. If such a voice is trained the quality will change to a brighter, more modulated, and less noisy one.

2.3. Androgen damages of female voices

This type of artificial dysphonia causes a lowering of the mean fundamental frequency, reduces the intonation range, and makes the voice sound rough and coarse like a man's voice. Observation of a glottal transverse insufficiency - just as in boys with mutation dysphonia - is characteristic of this disorder.

2.4. Vocal nodules

Bilateral benign tumours on the vocal folds are in most cases observed in children. In adult voices it most commonly is found in connection with hypercompression, - especially in the voices of singers and professional speakers.

The normal symptom is a gruff voice quality which is poorly equalized. Special difficulties arise during phonation in the middle register.

2.5. Chronical laryngitis

The symptoms of this group of voice disorders are mostly differentiated and cause difficulties when making the precise differential diagnosis.

In some cases the vocal folds are more or less oedematous and injected, now and then with hyperkeratosis and uneven edges. The voice quality changes very much and is often dominated by variations in compression, combined with pneumophonia and bad equalization.

2.6. Psychogenic dysphonia

In order to supplement the analysis with an examination of a characteristic non-organic type of dysphonia we have chosen the psychogenic dysphonia (functional dysphonia).

There is never any agreement between the laryngoscopic observation and the distinct changes in voice quality. It is typical for the psychogenic behaviour of these patients that they are not motivated for the clinical functional tests. It is therefore of great interest to include this group in the project. This means that the material thus enables us to compare the parameters from the normal voices with both organic and non-organic voice disorders.

3. Acoustic parameters for different voice qualities

As mentioned in 1.1. the auditive terminology of the different voice qualities is rather confusing. However, it seems that two or three main groups can be separated.

3.1. The weak, noisy, and breathy voice

At the physiological level this phonation is characterized by a more or less continuous airflow during the entire vibratory cycle, and no glottal closing phase can be observed. This is undoubtedly a result of either an insufficient medial compression of the arytenoid cartilages or a lateral excursion of the vocal folds.

At the acoustic level the escaping air generates noise especially at frequencies above 3000 Hz, and due to the lack of a glottal closing phase the acoustic vowel spectrum has very weak harmonics which reduce the levels of the higher formants. The first harmonic is very prominent.

The auditive impression is a breathy and noisy voice quality with a dull and dark timbre mainly caused by the lack of higher harmonics, but in some cases also caused by a lowering of the fundamental frequency. The voice is weak and the intelligibility is bad.

It seems quite natural that this voice quality appears particularly in paralyzed voices.

3.2. The harsh and rough voice quality

This voice quality has its origin in a disturbed vibration of the vocal folds. According to recent investigations (Wendahl 1963 and 1966, Isshiki et al. 1969, Lieberman 1963) the duration of the glottal vibrations changes from period to period even for a vowel phonated at a constant fundamental frequency. This irregularity of the vocal fold vibration is often due to excessive tension of the folds (Zemlin 1969). Furthermore, the harsh quality is combined with frequent use of glottal attack and often with a slightly lowered fundamental frequency.

The harsh or rough voice quality seems to dominate in voice disorders such as the different forms of vocal nodules and in some forms of laryngitis.

3.3. The hoarse voice quality

According to Isshiki et al. (1969) the auditively based term "hoarseness" comprises at least four auditive parameters, including breathiness and harshness.

Acoustically it manifests itself as a combination of noise and lack of harmonics in the upper part of the spectrum, and as a continuous fluctuation in periodicity (in extreme cases no pitch can be registered at all).

Physiological changes which occur (among other things) during the mutation, or changes caused by neoplasms or by vocal nodules and laryngitis, are the main causes for this change in vocal quality.

4. The temporal relations in pathologic speech

Apparently very little research has been made on the altered temporal relations in pathologic speech. It is a clinical experience that both the length of vowels and consonants and the length and number of the inspiratory pauses are extended in different voice disorders.

Normally the laryngeal paralyses will result in a greater consumption of air caused by the insufficient glottal closure. This means that the respiratory frequency is increased which in turn reduces the length of the voiced passages (phonation groups). An increased consumption of air may also be registered e.g. in laryngitis and often in psychogenic dysphonia. The function of the glottis when affected by recurrent paralysis will cause a lengthening of the transition time between the sound segments.

It is obvious that the changes in the temporal relations could be used as a tool for describing the different voice disorders, too. In this first trial we have therefore decided to use the length of the continuous voicing as a parameter (see 6.3.).

5. The most important acoustic parameters

5.1. The fundamental frequency

Information about the fundamental frequency may be obtained by an automatic scanning of the output voltage from a fundamental frequency detector. The collected data contain information about the lowest, the highest, and the mean fundamental frequency, the intonation range, the frequency and the range of falling/rising changes in the intonation, and the way in which these changes occur (e.g. continuously or in jumps).

If the automatic scanning is synchronized with the vocal fold vibrations, the parameter also contains information about variations in periodicity.

5.2. The relative intensity of the higher harmonics

An automatic scanning of the relative vowel intensity above 1000 Hz seems to be a good parameter for the amount of energy in the higher part of the spectrum. ("Relative" means relative in relation to the total energy at the moment of measurement.) A reduced energy of the harmonics in the F_2-F_3 region in relation to the F_1 region will result in reduced intelligibility. (I. Lehiste and G. Peterson 1959, S. Smith 1961).

The parameter contains information about the strongest and weakest relative intensity, the mean relative intensity, and the dispersion of relative energy above 1000 Hz for all the voiced sounds in a spoken text.

5.3. Duration of the phonation groups

We define a phonation group as a group of voiced segments, i.e. the phonation groups are interrupted by pauses and unvoiced consonants. As previously mentioned several phonation disorders are characterized by too short phonation groups caused by an insufficient closure of the vocal folds or lack of ability to close the folds. This parameter thus gives very valuable information about the ability to close the vocal folds and to keep the folds in the phonatory position. Indirectly, it gives information about the respiratory frequency.

Information derived from that parameter consists of measurements in milliseconds for the longest, the shortest, and the mean duration of the phonation groups together with details concerning the distribution of the durations in a spoken text.

6. How to extract the parameters

6.1. The fundamental frequency

A purely automatic scanning of the fundamental frequency will register a frequency of 0 Hz in all pauses which is of no interest in the further treatment of the data. Therefore the data collecting system should be stopped when the voicing stops. The data collection is under supervision of a "voice indicator" (see the next paper 3.1.) which starts the data collecting procedure 25 ms after the voicing starts, and stops it immediately when the voicing stops. The "voice indicator" is operated by the intensity level of the fundamental frequency. The speed of the data collection is set by a "pulse generator" (see next paper 3.2.) which is normally triggered by the mains supply at 50 Hz. The data are either stored on a punched paper tape (see next paper 3.3. and 3.4.) or fed on-line into a computer. The punched paper tape is an appropriate way of storing the data if the system is to be used in a phoniatric clinic without EDP-possibilities. The on-line processing calls for a small computer[3] connected with the data collecting system. In this way the EDP-output will appear immediately after the patient has been recorded, which is a great advantage in a routine analysis. By means of appropriate software the EDP-system will handle the data, and the output will primarily appear as histograms with standard deviations showing the distribution of the different fundamental frequencies in a spoken text.

6.2. The relative intensity of the higher frequencies

The relative measure of the intensity above 1000 Hz calls for two intensity registrations: (1) total intensity and (2) intensity above 1000 Hz.

---

3) Our laboratory has just ordered a computer, which will be used for on-line run of this project, as well as for other purposes.

The relative intensity above 1000 Hz in relation to the total spectral energy is expressed as:

$$\frac{I_{hp1000}}{I_{total}}$$

Electronically these calculations can be made by substracting the logarithmic value of the denominator from that of the numerator:

$$\log \frac{I_{hp1000}}{I_{total}} = \log I_{hp1000} - \log I_{total}$$

If, therefore, the two intensity measures are represented as logarithmic voltages (i.e. in dB), these voltages can be subtracted from each other in a differential amplifier, the output of which thus represents the relative intensity above 1000 Hz measured in dB.

The further treatment equals the treatment of the output voltage from the fundamental frequency meter (see 4.1. in this paper and 3.2. in the next paper). Ideally the fundamental frequency and the intensity should be registered simultaneously. However, due to the electronic circuitry there will be a short delay of 0.45 msec. between the registration of the two parameters (see Fig. 2 in the next paper). On the punched paper tape each fundamental frequency measurement will be followed by an intensity measurement.

The final representation in the EDP-output equals the representation of the fundamental frequency parameter, i.e. as histograms showing the distribution of the energy above 1000 Hz.

6.3. Duration of the phonation groups

The above-mentioned "voice indicator" registers when voicing is present, and in the "voice duration meter" (see next paper 3.3.) the duration of the phonation groups is converted to an electrical voltage, which is fed via the multiplexer and the A/D-converter to the puncher. The system then enables us to register 120 different durations (see next paper 2.3.).

On the punched tape the duration data are separated from the two other parameters by means of "boundary markers" generated by the "voice duration tape coder" (see next paper 3.4.).

The EDP-output represents the durations in the form of easily comparable histograms.

6.4. Resolution of the parameters

Considerations about the resolution of the parameters indicate that the registration of the fundamental frequency (6.1.) calls for about 100 decimal values per collected data and the resolution of the relative intensity registration (6.2.) requires about 60 values per data. The resolution of the time parameter (6.3.) depends very much upon the voice disorder considered, but again 100 values seem to satisfy our demands.

7. The value of non-acoustic tests

The present lack of empirical data concerning the acoustic changes in voice quality for different voice disorders justify a procedure which compares the above-mentioned acoustic parameters with the results of examinations made by the classical methods. We have planned to make complete diagnostic investigations at the University Hospital or at the Institute of Speech Disorders for all voices recorded, normal as well as pathological. In this connection it is our plan to make laryngoscopy and stroboscopy, to establish the total phonation range and the maximum duration of phonation, and according to requirements to supplement the results with the necessary examinations by means of X-ray, glottography, blood and hormon tests, etc. A profound anamnesis will be made of all patients in order to obtain the best possible classification. Only in this way it will be possible to get hold of and to get experience about the specific characterization of the acoustic parameters.

8. Some further acoustic parameters

This paper has been called "Some remarks ...". It is only meant as a description of a project which has just started. The three parameters which have been described in the paper seem, after pilot tests, to be the ones that contain the most relevant information. Naturally several other possibilities exist, and as the project proceeds we will test the utility of other parameters. Among these we only want to draw the attention to:

(1) the duration of the pauses in a spoken text,

(2) the relative intensity of the unvoiced consonants in relation to the vowels,

(3) separate analyses of the irregularities in periodicity, and

(4) correlation analyses of the intensity variations for the glottal cycles (Koike 1968).

References

Isshiki, Okamura, Tonabe, and Morimoto 1969: "Differential Diagnosis of Hoarseness", Folia Phoniatrica 21, 1.

Koike, Y. 1968: "Vowel Amplitude Modulations in Patients with Laryngeal Diseases", JASA 45,4, p. 839-844.

Lauritzen, K. and B. Frøkjær-Jensen 1970: "Comparative phonetic-acoustic analysis before and after speech therapy of voices suffering from recurrens paresis", ARIPUC 4, p. 145-165.

Lehiste, I. and G. Peterson 1959: "The Identification of Filtered Vowels", Phonetica 4, p. 161-177.

Lieberman, P. 1963: "Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathologic Larynges", JASA 35,3, p. 344-353.

Smith, Svend 1961: "On Artificial Voice Production", Proc. of the IV. Intern. Congr. of Phonetic Sciences, Helsinki, p. 96-110.

Sonninen, A.,and P.H. Damsté 1971: "An International Terminology in the Field of Logopedics and Phoniatrics", Folia Phoniatrica 23,1, p. 1-80.

Wendahl, R.W. 1963: "Laryngeal Analog Synthesis of Harsh Voice Quality", Folia Phoniatrica, 15, p. 241-250.

Wendahl, R.W. 1966: "Laryngeal Analog Synthesis of Jitter and Shimmer, Auditory Parameter of Harshness", Folia Phoniatrica 18,2, p. 98-108.

Zemlin, W.R. 1969: Speech and Hearing Science.