Speech waveform perturbation analysis revisited

Askenfelt, A. and Hammarberg, B.

journal: STL-QPSR
volume: 22
number: 4
year: 1981
pages: 049-068

http://www.speech.kth.se/qpsr
IV. TECHNICAL PHONIATRICS

A. SPEECH WAVEFORM PERTURBATION ANALYSIS REVISITED
A. Askenfelt & B. Hammarberg*
* Dept. of Logopedics and Phoniatrics, Huddinge University Hospital

Abstract

The clinical usefulness of seven acoustic measures of cycle-to-cycle variations (perturbations) in the speech waveform was investigated. Three of these measures referred to the frequency of occurrence and severeness of waveform perturbations in special selected parts of the speech identified by means of the rate of change of the fundamental frequency. Three other measures referred to statistical properties of the distribution of the relative frequency differences (DFo) between adjacent pitch periods. One perturbation measure referred to the percentage of consecutive pitch period differences with alternating signs. The acoustic measures were tested on tape recorded speech samples from 41 voice patients, before and after successful therapy. All measures were calculated automatically and applied to running speech. Scatterograms of the acoustic measures and a sum of perceived deviant voice qualities as rated by voice clinicians are presented. The different perturbation measures are compared as regards the acoustic-perceptual correlation and the ability to discriminate between normal and pathological voice status. The standard deviation of the DFo-distribution is suggested as the most useful acoustic measure of waveform perturbations for clinical applications.

Introduction

In a previous study an acoustic method for automatic detection and measurement of cycle-to-cycle variations (perturbations) in the speech waveform was described (Askenfelt & Hammarberg 1980). Three acoustic measures were presented which might serve as indicators of the frequency of occurrence and severeness of perturbations in the speech waveform. In the present study we have extended the repertoire of acoustic perturbation measures to seven, and also enlarged our patient group considerably. The aim of our work is still to find acoustic parameters which are relevant from a physiological point of view. As direct data on the physiological behavior of the voice organ are very laborious to obtain we have restricted ourselves to an estimation of the degree of vocal dysfunction by indirect means, i.e., perceptual evaluation by
voice clinicians. In our opinion clinical diagnosis is not a sufficiently accurate means of classification of voice function for the present purpose, since patients with the same diagnosis may differ greatly in voice quality. Perceptual voice ratings have earlier been demonstrated to be useful for acoustic aperiodicity estimation (e.g., Coleman 1971; Deal & Emanuel 1978). From an instructional film by Hirano (1976) it is also apparent that perceived voice qualities such as roughness and instability reflect irregularities of the vocal fold vibration.

A quantification of voice function will have its main application in speech clinics for evaluation of therapies as well as a support in diagnosis. The need for objective aids seems to be obvious to most experienced clinicians (c.f. Lubker 1979). It is also likely that a description of the vocal function in objective terms will lead to a better concordance in voice terminology.

Clinical considerations

In most earlier work on quantitative measures of speech waveform perturbations the methods of analysis have been applied on sustained vowels (Iwata & von Leden 1970; Hollien, Michel & Doherty 1973; Deal & Emanuel 1978; Hori 1979). Analysis of a short sentence has been performed by Lieberman (1963) and Hecker & Kreul (1971). The arguments for applying isolated sustained vowels in voice studies have been (1) accuracy of measurement, and (2) influence of variation due to voluntary speech patterns, such as phonetic context, stress and intonation (Iwata & von Leden 1970; Hollien et al 1973; Murry & Doherty 1980). However, vocal behavior is dynamic and voice quality variations due to pitch and loudness variations or onset and offset of voicing are certainly revealing as to voice function status and should not be disregarded in a thorough description of vocal dysfunction. For this reason a test with sustained vowels may not yield valid results. Our clinical experience is that the sustained vowel is representative of the patient's voice quality only in those cases when the voice is affected by a severe laryngeal pathology, such as laryngeal cancer, chronic laryngitis or unilateral paralysis. For the majority of the patients, however, it is necessary to analyze running speech in order to obtain a performance which is representative of the patient's voice function.
This problem of the representativity of the speech sample has not been discussed much in the literature. In a study on spectral noise and roughness relationships in male adults with laryngeal pathology, Hanson & Emanuel (1979) noted a substantial degree of relationship between vowel ratings and sentence ratings, but as the authors point out, not all the isolated vowels showed hoarse quality, although sentences from the same subjects were abnormally rough. As Horii (1979) points out, there is a paucity of data from large quantities of speech materials due to lack of efficient instrumentation and measurement procedures. As far as is known, only Gubrynowicz, Mikiel & Zarnecki (1980) have had access to computer techniques similar to ours, allowing the processing of large collections of data corresponding to several minutes of speech.

The aim of the present and earlier studies (Hammarberg, Fritzell, Gauffin, Sundberg & Wedin, 1980a,b) is to develop and apply such automatic procedures for acoustic analysis of running speech and thus increase our knowledge about the voice function.

Acoustic method

The technical equipment used for recording and fundamental frequency (Fo) tracking has been described in previous reports (Askenfelt & Sjölin 1980; Askenfelt & Hammarberg 1980). Tape recordings were made in a standardized manner. The patients read a tale of approximately 40 s duration in a sound treated booth. This speech sample corresponded to approximately 1800 and 3000 pitch periods for male and female voices, respectively. The speech was recorded on a two track audio tape recorder. The air-borne signal from a condenser microphone was recorded on one track and the signal from a contact microphone attached to the skin pretracheally above the sternum on the other. The recordings were processed by hardware equipment connected to a computer. The period time and sound intensity were sampled at the end of every pitch period and stored in the computer for later processing. The resolution of the frequency measurements was 1 Hz in the frequency range of interest (90 - 300 Hz), corresponding to a period time resolution between .12 and .01 ms.

The analyzing equipment was not capable of discriminating between pitch
perturbations ("jitter") and amplitude perturbations ("shimmer"). Both these kinds of cycle-to-cycle variations were regarded as pitch perturbations, which thus motivate the use of the neutral term waveform perturbation (WP) throughout this work. All WP-measures in this study were calculated automatically and applied to running speech.

Description of WP-measures

(1) WP-measures associated with the WP-selection strategy:
In our previous work (Askenfelt & Hammarberg 1980) three perturbation measures were used, Perturbation Factor (PF), Perturbation Magnitude (PM) and the product PF*PM. A thorough description of these measures can be found in that study. In short, the underlying strategy was to identify the WP-parts in running speech with the aid of the rate of change of fundamental frequency. After this was done the measures were calculated as follows.

* PERTURBATION FACTOR (PF): "How often does WP occur?"

$$PF = \frac{N_{wp}}{N_{tot}}$$

The number of WP periods ($N_{wp}$) compared to the total number ($N_{tot}$) of voiced pitch periods.

* PERTURBATION MAGNITUDE (PM): "How large is the mean perturbation per WP period?"

$$PM = \frac{1}{N_{wp}} \sum_{N_{wp}} \frac{|F_{n+1} - F_n|}{F_n}$$

The mean perturbation magnitude during WP-parts.

* PERTURBATION PRODUCT (PP): "How large is the mean perturbation per voiced period?"

$$PP = PF \times PM$$

The perturbation product (PP) will reach a high value if the perturbations are frequent (PP high) and/or if the magnitude of the perturbations are large (PM high). This influence of both occurrence and
severeness of the perturbations makes it natural to regard PP as a quality factor for the condition of the voice, a high value indicating an unhealthy voice. The perturbation product can also be interpreted as the mean perturbation magnitude per voiced period.

In contrast to the strategy for the measures just mentioned all voiced periods were taken into account when calculating the following measures.

(2) WP-measure associated with directional changes of Fo:
* DIRECTIONAL PERTURBATION FACTOR (DPF): "How often does the Fo-curve change direction from period to period?"

$$\text{DPF} = \frac{N_{+/-}}{N_{tot} - 1}$$

The percentage of the total number of differences between adjacent pitch periods ($N_{tot}$) for which there is a change in the algebraic sign ($N_{+/-}$), see Fig. IV-A-1. This measure was first proposed by Hecker & Kreul (1971).

![Diagram](image)

**Fig. IV-A-1.** Explanation of the Directional Perturbation Factor (DPF). DPF is calculated as the ratio between the number of adjacent pitch period differences with alternating signs (as in the segment in the figure marked Z) and the total number of pitch period differences.
(3) WP-measures associated with the DFo-distribution:
The following WP-measures apply to the distribution of the relative frequency differences (DFo) between consecutive pitch periods. Referring to Fig. IV-A-1, DFo was calculated as:

\[ DF_o = \frac{F_{n+1} - F_n}{F_n} \]

Statistical properties of DFo-distributions have been used earlier as successful WP-indicators (Gubrynowicz et al 1980). Already Lieberman (1963), the pioneer in the field of waveform perturbations, used a distribution of pitch differences, although not normalized, as a basis for his WP-measure. The DFo-distributions can be expected to show different shapes for normal and pathological voices, respectively, (Liebermann 1963, Gubrynowicz et al 1980), as indicated in Fig. IV-A-2. Pathological voices tend to have proportionally greater amount of perturbations with large magnitude.

Fig. IV-A-2. Distribution of the relative frequency Differences between adjacent pitch periods (DFo) for a female patient (unilateral papilloma) before (solid) and after (dotted) successful treatment.
* DELTA-Fo DISTRIBUTION WIDTH (DFoW):
The dispersion in the DFo-distribution was expressed by the standard deviation, henceforth DFoW. Voices with abnormal function would be expected to have higher DFoW values than normal voices.

* DELTA-Fo DISTRIBUTION ZERO (DFoZ):
The DFo-distributions for normal voices often showed a pronounced peak at DFo=0. Due to truncations in the computations this peak corresponded to DFo-values between 0 and ±1%. It was hypothesized that it would be advantageous to include a detailed description of this characteristic in the shape of the distribution as a separate WP-measure.

* DELTA-Fo DISTRIBUTION PEAK (DFoP):
The height of the DFo-distribution at origo (DFoZ) compared with the standard deviation of the distribution (DFoW).

\[ \text{DFoP} = \frac{\text{DFoZ}}{\text{DFoW}} \]

Because of the curious shape of the DFo-distributions with an extreme peak at origo, Fig. IV-A-2, it seemed worthwhile to test if a combination of DFoW and DFoP could describe the shape of the distribution better than any of these measures alone. As mentioned earlier, a healthy voice will have a comparatively narrow DFo-distribution with a high peak at origo and a small standard deviation. Consequently the DFoP-value associated with a normal voice will be high. A pathological voice can be expected to show the opposite characteristics.

Perceptual ratings

In order to test the usefulness of the acoustic measures, perceptual voice analysis by experienced voice clinicians was applied. The judges were logopedists and phoniatricians at the Dept. of Logopedics and Phoniatrics, Huddinge Univ. Hospital. Since 1977 they had participated in a systematic training and dialogue in matters of perceptual voice evaluation by assisting in perceptual voice evaluation experiments of the same kind as in the present study (Hammarberg et al 1980 a,b;
Hamarberg, Fritzell & Schiratzki 1981). In these studies the perceptual characteristics of pathological voices were quite accurately classified, as reflected by the high intrajudge reliability between .93 and .97 (Hamarberg et al 1980a).

The voice sample
In the previous study Askenfelt & Hammarberg (1980) perceived deviance from normal voice quality, rated by clinically experienced logopedists in terms such as roughness, diplophonia, vocal fry and unstable voice quality, showed high correlation with a measure of waveform perturbation (FP). In order to test the new acoustic perturbation measures a new and larger voice sample was analyzed. Tape recordings before and after therapy from 41 patients (11 male, 30 female), were selected from a sample of 135 voice patients according to the following three criteria:

(1) tape recordings before and after therapy had been made under identical conditions,

(2) the pre therapy recordings were characterized by a high degree of roughness, diplophonia, vocal fry and/or unstable voice quality, and

(3) the post therapy recordings showed perceived improvement of voice quality as compared to the pre therapy recordings, i.e., the post therapy recording could serve as a 'control' as it sounded normal or almost normal.

The eleven male patients ranged in age from 23 – 60 years with a mean age of 38 years. The 30 female patients ranged from 22 – 75 years with a mean age of 42.5 years. The bias in favor of women in the voice sample reflects the predominance of female voice patients in speech clinics. Laryngologic examination of the 41 voice patients – described in the patients' records – was in most cases done routinely by indirect laryngoscopy, in some patients complemented with photography, fiberoptic laryngoscopy or stroboscopy. Recordings of the vibratory pattern of the vocal folds were not made. The diagnoses represented in the voice sample, Table I, were organic disorders in 33 of the 41 patients with a predominance of mass lesions of the vocal folds. As for the eight patients with functional dysphonia, no abnormalities of the vocal folds were found, except for one patient with a slight asymmetry of the vocal fold vibrations as found during stroboscopic examination.
Diagnoses

<table>
<thead>
<tr>
<th></th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unilateral polyp</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Bilateral polypoid thickening</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Unilateral vocal nodule</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Bilateral vocal nodules</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Laryngeal carcinoma (T1 A)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Chronic laryngitis</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Acute laryngitis</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Papilloma (unilateral)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Unilateral paralysis</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Functional</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional dysphonia</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>11</td>
<td>30</td>
</tr>
</tbody>
</table>

Table I. Diagnoses represented in the voice sample.

The predominance of mass lesions in this material, primarily chosen on the roughness-instability criteria, is in accordance with findings by Iwata & von Leden (1970), Hanson & Emanuel (1979) and Ludlow, Coulter & Gentges (1981).

At the time of the post therapy recording the laryngologic examination of the organically disordered patients showed a normalized or markedly improved laryngeal condition due to surgery or/and voice therapy. Among the five patients with paralytic dysphonia, three were improved by Teflon injection in combination with voice therapy, and two by voice therapy alone.

Listener tasks

The pre and post therapy tape recordings of each of the 41 patients were randomized. In a pilot evaluation session three logopedists were asked to evaluate 30 of the recordings by describing the voice qualities in their own terms. From these evaluations 12 voice describing terms were selected, Fig IV-A-3. Six logopedists of the jury, earlier mentioned, rated the tape recordings, male and female voices separately. Thus, 82 recordings were evaluated in four different sessions. The stimulus tapes
were played from loudspeakers in a sound treated room. For each recording, voice quality was rated on a 5-point scale for each of the ten voice describing terms. A rating of 4 represented a severe degree of abnormality and 0 the absence of abnormality. Voice register and pitch were rated on separate scales. The judges listened to each recording five times. For the evaluation of the acoustic measures a sum of the mean ratings of seven voice describing terms was used, vocal fry/creaky, diplophonia, voice breaks, grating/harsh, rough/course, instability and register. This summed perceptual rating will henceforth be referred to as PERC7.

<table>
<thead>
<tr>
<th>Voice no</th>
<th>Initials of judge</th>
<th>Date</th>
<th>Voice quality to be rated</th>
<th>none</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>1 Breathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>2 Vocal fry/creaky</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Sw. knarr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>3 Hard glottal attacks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>4 Hyperfunctional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>5 Hypofunctional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>6 Diplophonia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>7 Voice breaks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>8 Grating/harsh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Sw. skrep)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>9 Rough/course</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Sw. skrovlig,grov)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>10 Instability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>11 Register</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>chest/modal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>middle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>head/falsetto</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>12 Pitch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rather low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rather high</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>high</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. IV-A-3. Perceptual voice evaluation sheet.
Results and discussion
The set of seven different WP-measures were computed for all 41 patients pre and post therapy. The results are visualized in the 14 plots shown in Fig. IV-A-4 (male voices) and Fig. IV-A-5 (female voices). In these figures the WP-measures are plotted versus the summed perceptual rating PERC7 earlier described. The correlation coefficient (r) and a regression line are also included in the plots. Data for individual patients pre and post therapy are indicated with triangles and dots, respectively, and connected with straight lines.

It is evident from these plots that there is a connection between perceived voice quality and nearly all of the acoustic WP-measures. However, the information power, reliability and usefulness for clinical purposes vary substantially between the WP-measures as will be discussed later.

It should be stressed that a one-to-one relationship between an acoustic measure and rated voice quality never is to be expected. Of course, this is not a characteristic unique to acoustic parameterization of human behavior but can be found in all cases in which a measurable physical parameter is used for estimating the condition of a human function, e.g., blood pressure as an estimator of health status. Despite of this lack of strictly mathematical correspondence between the measured physical parameter (WP) and perceptually rated voice function (PERC7), acoustic parameters are likely to prove useful as an aid in voice diagnosis and treatment, in the same way as information on the blood pressure has been found useful in everyday medical work.

As mentioned in the introduction we are trying to develop methods which can be applied in speech clinics. Thus, the demands on the methods of measurement are not only restricted to expose important aspects of voice function. The acoustic measures must also be reliable in order to meet the demands in clinical work. In other words, the measures should carry essential information on the actual voice status at the same time as they are immune to irrelevant signal properties. The characteristics of the seven WP-measures will now be discussed individually with the above comments in mind. References will be made to Fig. IV-A-6 in which the WP-measures are compared, male and female voices separately. The leftmost diagrams in this figure show the correlation coefficients. The diagrams in the middle show a discrimination factor, calculated as the percentage change along the regression line in the WP-values between pathological (PERC7=100) and normal (PERC7=0) voice quality.
Fig. IV-A-4. Scatterogram of the different WP-measures and the summed perceptual rating PERC7 (see text), male voices. Data for individual patients' pre and post therapy are indicated with triangles and dots, respectively. Data beyond the range of the plots are indicated with a jagged line together with the corresponding value. The correlation coefficient, $r$, the repression line (dashed) and a discrimination factor DF (see text), are included in each scatterogram.
Fig. IV-A-5. Scatterogram of the different WP-measures and the summed perceptual rating PERC7 (see text), female voices. Data for individual patients pre and post therapy are indicated with triangles and dots, respectively. Data beyond the range of the plots are indicated with a jagged line together with the corresponding value. The correlation coefficient $r$, the regression line (dashed) and a discrimination factor DF (see text) are included in each scatterogram.
A high DF-value means that there is a large percentage difference in the WP-value between abnormal and normal voice function, respectively. The rightmost diagrams in Fig. IV-A-6 show the product r*DF, which will be commented on below.

Fig. IV-A-6. Comparison of the seven WP-measures for male voices (top) and female voices (bottom), respectively. The figure shows the magnitude of the correlation coefficient r (left), a discrimination factor DF (middle), and the product r*DF (right), see text. The hatched portions in the leftmost diagrams indicate 95% confidence intervals for the correlation coefficient.

Discussion of the individual WP-parameters

*Perturbation Factor (PF), representing the occurrence of WP in terms of the percentage of WP-periods in the speech sample, showed a reasonably high correlation with the perceptual rating for both male and female voices. Further, it had the highest discrimination factor of all WP-measures for female voices and the second highest for male voices. This was a surprisingly good result as will be commented on below in connection with PP.

*Perturbation Magnitude (PM), representing the severity of WP in terms of the mean magnitude of the perturbations occurring during the WP-parts, was found to be the most dispersed of all measures. Probably this is because the numbers of WP-periods are mostly rather low, thus making the mean value sensitive to singular extreme samples. The correlation coefficient is close to zero and the measure tells almost nothing about the condition of the voice.
* Perturbation Product (PP), the product of PF and PM, representing the product of WP occurrence and severeness respectively, showed higher correlation coefficient and discrimination factor than PF for male voices but not for female voices. It was not expected to find the PF-measure superior to PP for any of the groups as PF only takes the occurrence of WP into account in contrast to PP which reflects both occurrence and magnitude of the perturbations. The explanation is probably that the magnitude of the perturbations (PM) are more or less random and thus it is mainly the frequency of occurrence (PF) which gives significant contributions to PP.

* Directional Perturbation Factor (DPF), i.e., the amount of adjacent pitch period differences having alternating signs, showed the highest correlation coefficient of all WP-measures tried as regards the male voices. Unfortunately the DF-values were low for both male and female voices which means that information on voice status associated with this measure is poor. The reason for this would be that the numeric value of DPF is as high as 40% for normal voices. This value should be compared with the highest DPF-value in the entire material, 75%, which emanated from the most extreme voice. These data yield a small max/min ratio and prevents conclusions regarding voice function from a given DPF-value. Still Hecker & Kreul (1971) reported that they were able to discriminate between five male patients with laryngeal cancer and five normal controls with the use of DPF. Their DPF-values were generally somewhat lower than those found in the present investigation.

* Delta-Fo Distribution Width (DFoW), i.e., the standard deviation of the DFo-distribution, showed both a high correlation with the perceptual ratings and also a rather high difference between normal and abnormal voices. The values for post therapy voices rated as normal were seldom higher than 5% which is in perfect agreement with the results of Gubrynowicz et al (1980).

* Delta Fo-distribution Zero (DFoZ), i.e., the percentage amount of pitch period differences with \(|\text{DFo}| < 1\%\) was less informative on voice quality than the standard deviation, i.e., DFoW.

* Delta Fo-distribution Peak (DFoP), i.e. DFoZ divided by DFoW, showed the same high correlation coefficients as DFoW and DFoZ. The
discrimination efficiency was higher than for DFoZ but lower than for DFoW. This observation supports the hypothesis earlier mentioned that it is the amount of large perturbations which should be emphasized in an acoustic measure of perturbations. This aspect of the perturbations is most clearly expressed by DFoW.

Ranking of the WP-measures

The product \( r \cdot DF \) was used as a basis for a ranking of the WP-measures. The usefulness of a WP-value for clinical purposes can be inferred from the \( r \cdot DF \)-product in the following way, c.f. Fig. IV-A-6.

A high correlation coefficient implies that the data points in the scatterogram are clustered close to the regression line, i.e., there is a close relationship between perceptual ratings and magnitude of the WP-measure. This means that the acoustic measure gives a reliable estimation of a patient's rated voice quality, as a certain change in the perceptual rating always will give rise to a proportional change in the WP-measure. This change in rated voice quality is also supposed to reflect a corresponding change in voice function, according to the reasoning in the introduction.

Further, a high discrimination factor means that the regression line has a steep slope, i.e., there is a considerable percentage difference between the WP-values for normal and pathological voices respectively. This is another necessary characteristic of an efficient WP-measure as there always will be certain variations in WP-values between patients with similar voice quality. A high discrimination factor reduces the influence from inevitable variations in the WP-values on estimated perceptual rating. In this way a safe discrimination between normal and deviant voice function with the aid of acoustic measures is assured.

Consequently, only WP-measures with both high correlation coefficients and high discrimination factors, i.e., high \( r \cdot DF \)-products, are promising candidates for a clinical applicable aperiodicity measure.

Using this criterion the PP- and PF-measures are the optimal ones of the WP-measures examined. The second best candidate would be DFoW. However, a closer inspection of the data plots in Fig. IV-A-4 and Fig. IV-A-5 reveals that PP and PF for both male and female voices gain considerably in the correlation coefficient as well as in the discrimination factor.
due to singular peripheral data in the plots. This is not the case with DFOW which shows a reasonable scatter of data around the entire regression line.

Furthermore, the magnitude of the PP- and PF-values turned out to be very sensitive to any changes in the analysis equipment. This is explained by the following fact. The selection of WP-parts out of the recorded speech which underlies the PF- and PP-calculations is based on the time derivative of the fundamental frequency (Askenfelt & Hammarberg 1980). Any changes and/or errors in the fundamental frequency measurement may cause dramatic changes in the time derivative which probably will influence the selection of the WP-parts seriously in most cases. Consequently, a considerable change in the PF- and PP-values will result.

The performance of DFOW was completely different. As DFOW is a measure of the dispersion of data in a distribution, a moderate number of changes in data will not influence the numeric value of the dispersion significantly. This property makes DFOW a comparatively robust measure, essentially independent of details in the analysis of the fundamental frequency.

The above considerations speak in favor of DFOW rather than PP and PF as the most suitable of the seven waveform perturbation measures examined as regards clinical applications. This conclusion is in agreement with the results from the large study by Gubrynowicz et al (1980), in which voice samples from 143 persons were examined by use of computer techniques. The most successful of their aperiodicity estimators was a measure of the dispersion in the DFO-distribution very similar to DFOW.

Concluding remarks

Much work has been devoted to the problem of finding acoustic correlates to voice dysfunctions. As mentioned above, most of the work has been restricted to sustained vowels or short sentences, mainly due to the extremely tedious manual processing of the registered waveforms. The automatic methods of the type used in this investigation are, of course, advantageous because of their uniform processing of data together with
the high processing speed and accuracy. One may object that they are not necessarily as reliable as the manual methods in identifying the individual pitch periods, especially in the cases of large perturbations where the consecutive periods deviate considerably. In other words, the accuracy of the automatic methods is high until an ambiguity in the quasi-periodic pattern of the speech waveform occurs. The errors which thus may arise will probably be of serious nature. On the other hand, it can be argued that the automatic methods are capable of processing so many more samples than the manual methods, that such errors will not have any significant influence on the result. This is probably a correct conclusion but in the case of extreme pathological voices the problem should perhaps not be neglected.

In a study of variations in the speech waveform it seems natural to devote a careful attention to the question of fidelity in the recording and reproducing of the waveform. This aspect has not been seriously considered in this study. The voices were recorded with an ordinary audio tape recorder and replayed at the time of analysis with another audio tape recorder. This procedure leads to an uncontrollable amount of phase distortion in the signal analyzed, which will influence the WP-measurements to an unpredictable degree. The phase distortion problem can, of course, be circumvented by the use of an FM-tape recorder. However, as the perceived "rough" quality of the voices due to waveform perturbations is present in the phase distorted replayed signal and thus still accessible for analysis, it seems possible to avoid the expensive FM-recording technique. This aspect is not the least important if the acoustic analysis techniques are to be implemented in the speech clinics.

As a substitute to the FM-recordings, it is possible to use a phase compensating device at the time of analysis (Holmes 1975). Another solution, used in the present study, is to use the identical or equivalent audio tape recorders throughout. This would guarantee the same phase distortion on all recordings and hence the same amount of bias in all analysis results. If none of these methods is practised the control over the phase relations in the signal is lost and, as a consequence, the reproducibility of the analysis results will be poor.

The question of whether it is pitch perturbations (jitter) or amplitude perturbations (shimmer) which are the main contributions to the WP-
content in the voices is not possible to study with the use of the present method. The ratio between jitter and shimmer is probably different for most patients, depending on the patient's type of voice dysfunction. Some authors claim that shimmer contributes essentially more than jitter to the perceived degree of voice roughness (Deal & Emanuel 1978, Nichols 1979). Variations in amplitude between adjacent periods would thus be far more serious to voice quality than variations in the period times. Nevertheless, it can be argued that the sensitivity of the analysis equipment to jitter and shimmer respectively is irrelevant for clinical purposes, if it can be shown that measured perturbation values reflect voice quality.

Finally, a last remark on the methods. Little information on the physiological causes of the aperiodic performance of the voice organ can be obtained from acoustic measurements of the aperiodicity in the speech wave alone, regardless of the sophistication level of the methods involved. Such basic data can only be collected from simultaneous acoustic recordings and registrations of the motion of the vocal folds, e.g. by means of high speed photography. The need for such physiological data is emphasized by most authors on this subject. The value of acoustic correlates to voice quality is the evident immediate clinical application earlier mentioned, which motivates the wealth of contributions to this field, including this present study.

Acknowledgements

The authors are indebted to Johan Sundberg for encouraging advice during the work on this project and for valuable comments on the manuscript. This project was supported in part by The Swedish Medical Research Council, project No: B82-17X-04783-07.

References


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


