Variation of voice quality features and aspects of voice training in males and females
Sulter, Arend Marten

IMPORTANT NOTE: You are advised to consult the publisher’s version (publisher’s PDF) if you wish to cite from it. Please check the document version below.

*Document Version*
Publisher's PDF, also known as Version of record

*Publication date:*
1996

*Link to publication in University of Groningen/UMCG research database*

*Citation for published version (APA):*
A structured approach to voice range profile (phonetogram) analysis

Arend M. Sulter, Hero P. Wit, Harm K. Schutte* and Donald G. Miller*

Department of Otorhinolaryngology, University Hospital Groningen and *Voice Research Lab, Department of Medical Physiology, Faculty of Medicine, University of Groningen, Groningen, the Netherlands


INTRODUCTION

Phonetography is a practicable and readily accessible method to investigate and map the quantitative potentialities of vocal output (1-4). The maximally loud and soft phonations throughout the entire frequency range are indicated in a plot of frequency against sound pressure level (SPL).

Figure 1 gives an example of a normal phonetogram\(^1\) from a male subject without vocal complaints. The datapoints in the plot are acquired during a short session in which the investigator asks the subject to phonate as loudly and softly as possible at selected frequencies,\(^2\) thereby covering the subject's whole

\(^1\)Many synonyms of the graphical representation of an individual's voice potentialities are proposed in articles concerning phonetography. The terms phonetogram, phonogram, voice range profile, voice field, voice area and F\(_{\sigma}\)-SPL profile have been used in literature. In this article the term phonetogram is used.

\(^2\)The frequency range was sampled at four pitches per octave, e.g. C3-E3-G3-A3 (in this octave: 131, 165, 196, 220 Hz).
measuring device have been incorporated in a computer (7), reducing the time required for both the acquisition of data and the graphical conversion into a phonetogram. This makes a visual feedback of the measurements possible for both subject and investigator. A further contribution to the automatic registration of phonetograms has been the incorporation of a unit into the equipment to determine fundamental frequency (7-9). The benefit of this unit is twofold: subjects or patients not able to sustain the given pitch can use an alternative (freely chosen) pitch. In addition, the occurrence of octave-errors and other mistakes in determining the correct pitch (which are already small when the registration is performed by experienced investigators) will be minimal. The computer also makes it possible to create immediately processable phonetographic data files.

After the first description of phonetogram-like profiles by Wolf, Stanley, & Sette (10) and an early article by Calvet & Malhiac (1), the method received sporadic attention in the literature (4,5,11-14). In recent years, however, a growing number of practical and theoretical articles on vocal function and voice use have dealt with phonetography. Recommendations were formulated to standardize procedures in the acquisition of phonetograms (6,15); the potential of phonetography as a clinical tool was illustrated (3,16,17); and the theoretical bases of profiles were questioned (2,18-20). The practical uses of phonetography, as reflected in the literature, can be summarized as: (a) assessing information about individual voice potentialities, (b) investigating the influence of therapy or surgical intervention and (c)
comparing phonetograms of selected groups (11,21-23).

The lack of clear parameters applying to the phonetogram as a whole, however, presents an obstacle in the comparison of one phonetogram with another, as well as in the establishment of standard reference values for the phonetogram.

Approaches in dealing with this problem are based on averaging (10), or rescaling techniques. With the latter technique the individual phonetogram is rescaled with the x-axis (frequency range) to 100% (3,11,21,23). After a number of phonetograms have been normalized in this way, summary statistics on intensities of vocal output can be compiled. Frequency-dependent intensity information, however, cannot be derived from these statistics.

In another approach Klingholz and Martin (2) have attempted to describe mathematically (half axis, vertices and rotation) an arbitrary number of ellipses that can be fitted on to phonetograms. However, the number of ellipses contained in a phonetogram is not specified, and there are various ways of fitting an ellipse through datapoints. Also, the acquisition of datapoints introduces an unpredictable deviation from the ideal ellipse shape. This lack of a consistent basis for analyzing the phonetogram with ellipses calls into question the validity of the results.

A different approach toward the analysis of individual phonetograms is proposed in this research note. Parameters representing three expert-acknowledged features are extracted from phonetograms. Advantages of this method include (a) the derivation of features from phonetograms without distorting its shape, and (b) the particular attention paid to the dynamic possibilities of the F₀-SPL range used in normal speech. To demonstrate this method of automated evaluation, a normal male phonetogram as well as a pathologic male phonetogram are processed, and the resulting parameter values are compared with normative male data. A future article will present these normative data and data of groups of subjects that have received vocal training over a period of at least two years (24).

METHODS

Features of phonetograms

A group of four speech therapists and three Ear, Nose, and Throat (ENT)-physicians were informally asked to describe the way they visually analyzed phonetograms and to give their opinion about what features should be regarded as important characteristics. The descriptions offered by this group included three common features:

Shape. The experts considered the shape a very important feature. The general shape of the phonetogram is complex, but it can be seen as the sum of
two overlapping ellipses, each with a different slope of the long axis (2). The intersection where the two ellipses meet in the loudest phonation contour is a typical characteristic of the phonetogram of subjects without voice training. In that specific place, in male subjects at about G4 (392 Hz) and in phonetograms of women slightly higher, at about A4 (440 Hz), a local minimum can be seen (see Figure 1). This local dip can be attributed to the transition from chest to falsetto register when the phonetographic datapoints are measured for the vowel /a/ (18). This interruption in the otherwise rising contour of maximum SPL is minimized by vocal training (5,11).

Enclosed area. Connection of the lines of the loud phonation contour with the soft phonation contour (the upper and lower part of the phonetogram respectively) yields an enclosed area. All observations and judgments of phonetograms take this area into account. However, lack of quantitative knowledge about what constitutes a "normal" area results in a qualitative judgment with an imaginary frame of reference. The same can be said about the frequency range: a minimum of two octaves is often used in practice (3,5,25). However, only limited knowledge is available concerning the mean range and standard deviation of the frequency range in large specified groups of men and women.

"Speaking Range" dynamics While the phonetogram covers the entire frequency range, the speaking voice in its normal function uses only a part of the range. In order to reflect the importance of this portion of the range, a formula was devised to analyze it with respect to mean speaking fundamental frequency (mff).

Parameters describing the features

Representative parameters can be defined to describe in an approximate way the different features (shape, enclosed area and "speaking range" dynamics). A relatively large number of parameters (for the feature shape) are introduced in order to increase the chance of detecting deviations from a normal pattern. A second purpose is to promote the emergence of constellations of parameter values specific to pathologic entities. Because of the large number of parameters, however, considerable redundancy can be expected.

---

3The x-axis of a phonetogram extends from C₁ to c⁵ in the Helmholtz notation or from C1 (32.7 Hz) to C8 (2096 Hz) in American notation (6).
Shape-related parameters: Fourier Descriptors. Rather than normalizing the frequency range or mathematically describing arbitrarily projected ellipses in the phonetogram, the shape itself is analyzed with Fourier Descriptors (FD). Fourier Descriptors were developed for computerized reading of handwritten alphanumeric characters (26). In this procedure a closed contour, consisting of straight line segments, is transformed into a set of slope values as a function of length along the contour. Starting the analysis at a certain point on the perimeter (in our case the point corresponding to the lowest loud phonation frequency), one proceeds clockwise in steps (see Figure 2a). Each step consists of the computation of the length of a line segment and the angle between this segment and the following one. This procedure results in a set of length values and a set of angle values, giving angle as a discrete function of length along the contour (see Figure 2b). Formulas were developed by Zahn and Roskies (26) to calculate the Fourier transform of this function, taking into account that the points along the length-axis of the function (taken as the independent variable) are in general not equally spaced.

Figure 2. The shape of the phonetogram is analyzed with Fourier Descriptors. a) Starting at the lowest loud phonation, lines are drawn between the phonetogram points in a clockwise direction (l, l₁, l₂, ...). Next, the angle between adjacent lines is calculated (N, N₁, N₂). b) Line lengths and angles are placed in a plot with new axes. c) The information in the plot with the length of line segments and angle axes is processed with a Fourier transform, resulting in a number of Fourier Descriptors. Close to the origin the general shape is defined, whereas the Fourier Descriptors higher on the x-axis represent small changes in shape. The amplitude of an FD gives its relative contribution to the shape.
Section 7

As a first attempt to investigate the usefulness of this shape quantification procedure, amplitude values of the calculated Fourier transform are displayed. These amplitude values are called "Fourier Descriptors". In the plot of Figure 2c on the x-axis a discrete number of thirty Fourier Descriptors are given. The lowest numbers define the general shape, whereas the FDs higher on the x-axis represent small changes in shape. The y-axis gives the amplitude of each FD representing its relative contribution to the shape. As an example one can consider the FDs for a circle and an ellipse: For the sake of simplicity, the length intervals with which the contour of the circle or ellipse is sampled are assumed to be equal. In the case of a circle, angle as a function of length is constant, resulting in a value of zero for all FDs. Following the contour of an ellipse, angle as a function of length will have two maxima (for the "sharp ends" of the ellipse) and two minima (the long sides of the ellipse). The magnitude of this function gives the value of FD. In general FD is a measure for the ellipticity of a contour. When the angle function shows three maxima and minima in tracing the contour, this will be reflected in the magnitude of FD, and so forth.

Contour regularity. Even when care is taken for a proper acquisition of phonetogram points by following UEP procedures (6), in many cases the perimeter (especially along the loud phonation contour) has an irregular aspect (20). The parameter which illuminates this aspect is the contour regularity. This ratio is derived by dividing the enclosed area of the phonetogram by the squared perimeter, yielding a dimensionless figure. The highest contour regularity value will be derived from a circle, with greater irregularities yielding smaller values. Deviations from the circular shape correspond to smaller contour regularity values.

Phonetogram slope. A central straight line is drawn through the phonetogram. The slope and position of the line are determined by the least possible sum of squared distances to the measured points. Figure 3 shows the procedure determining the position and slope of the central line through the
Phonetogram Analysis

Figure 4. The enclosed area is calculated in four steps. a) The integral of the loud phonation contour is calculated. b) The integral of the soft phonation contour is determined. c) A subtraction gives the enclosed area. This area is divided by the area of the outlined rectangle. This yields a dimensionless figure relating to the part of the phonetogram covered by a subject's phonatory capabilities.

Area-related parameters. Enclosed area. The phonetogram is separated in a loud phonation contour (upper part) and a soft phonation contour (lower part). In case there is only one measuring point at either the lowest or the highest produced frequency, this point is regarded as belonging to both the soft and loud contour. After computing the area between the lower contour and frequency axis, this area is subtracted from the area between the higher contour and the frequency axis (see Figure 4a and 4b). The remaining enclosed area is divided by a constant, namely, the area of the rectangle with corners 32.7 Hz - 40 dB, 32.7 Hz - 110 dB, 2096 Hz - 110 dB and 2096 Hz - 40 dB (see Figure 4c). The reference area is based on axes proposed in the recommendations of the UEP (6). We chose a rectangle with the y-axis from 40 to 110 dB because in clinical practice measured vocal loudness hardly exceeds the intensity level of 110 dB, and this range has a center intensity of 75 dB. This reference line is hereafter employed in the analysis of the "Speaking Range" dynamics. By using a relative quotient, this quotient is thus independent of the scaling of x- and y-axes. Furthermore, this dimensionless ratio was chosen to increase comprehensibility: The size of determined enclosed area can be directly related to the frame of the phonetogram described above. For instance, a value of 0.238, as in figure 4, indicates that almost a quarter of the reference rectangle is covered with the
phonetogram area.
Frequency range.

\[ x_{\text{oct}} = \frac{6 \times (\log (x_{\text{Hz high}}) - \log (x_{\text{Hz low}}))}{\log 2096 - \log 32.7} \]  

(1)

With equation (1) the individual frequency range can be obtained as a number of octaves \( x_{\text{oct}} \) after the highest \( x_{\text{Hz high}} \) and lowest \( x_{\text{Hz low}} \) possible phonatory frequencies have been determined and the difference between these frequencies is calculated. Figure 5 illustrates this procedure.

Parameters related to "Speaking Range" dynamics. At four selected frequencies the dynamic range and the central position of this range are determined. These data provide information about the capacities for a person to modulate frequency and intensity within an arbitrarily determined speaking range. Two ways of selecting these frequencies are proposed here.

The first one is independent of the absolute frequency scale and uses information on the individual mean speaking fundamental frequency (mff). In the other procedure standardized male and female mff's are used. The mff of male subjects was set at 123 Hz, while a mff of 220 Hz was chosen for female subjects (see also Awan, (27), using mff's of 123 Hz and 206.6 for male and female subjects, respectively). In both procedures the other three frequencies at which the dynamic range is investigated are: three semitones below mff, half an octave and an octave above mff. We assumed that with these frequencies the speaking voice range is largely covered.

With a microphone at a distance of 30 cm from the mouth, measured mean intensities of normal speech will generally fall between 60 and 80 dB (27). This intensity range is therefore most important for a normal production and

Figure 5. The lowest (\( F_{\text{minimum}} \)) and highest phonation (\( F_{\text{maximum}} \) ) are transposed on a tone scale. The difference between the extremes gives the frequency range in number of octaves.
Phonetogram Analysis

Figure 6. To assess "Speaking Range" dynamics at four frequencies (A, B, C and D) the distance from the reference line of 75 dB to the loud phonation contour (a) and the soft phonation contour (a) is determined. The four frequencies are: the mean speaking fundamental frequency (mff) (A), mff minus 3 semitones (B), mff plus half an octave (C) and mff plus an octave (D). Because of the relative importance in normal speech of intensities around 75 dB the distances are processed with a weighting factor. In the figure the weighting factor is indicated in the third dimension. With the distances from the reference line and the imposed weighting factor the center of the dynamic range can be determined.

communication of running speech. A restricted intensity range may affect intonation and stress patterns and thus reduce the quality of spoken language. The intensities above 80 dB and below 60 dB normally will not be used during speech; however, the ability to raise one's voice is necessary for adapting to these special occasions which demand high intensities.

Because all intensities are not equally important during normal speech production a weighting factor was introduced in calculating the intensity range at the given frequencies. The weighting factor uses the natural logarithm of the measured values. It enhances the importance of intensities used in normal speech, in contrast to extremes in vocal loudness that are only used occasionally in shouting (loud voice) or quiet conversation (soft voice). The line representing an intensity of 75 dB is arbitrarily selected as the reference intensity for a normal intensity modulation. On both sides of the 75 dB line the importance of the intensity decreases approximating the decay of a natural logarithm. Figure 6 gives the selected frequencies together with a graphical illustration of the weighting factor.
Weighted dynamic range. At the four frequencies the distances (in dB) of the measured minimum and maximum intensity from 75 dB are calculated. When a minimum value of, for instance, 55 dB is measured, the distance from the reference line is 20 dB. This relative value is processed with the weighting factor; that is, the natural logarithm is taken of the value, resulting in a new weighted value of 3.0. The same procedure applied to a maximum intensity of, for instance, 85 dB results in a weighted value of 2.3. The weighted dynamic range (55 - 85 dB) thus gives a value of 3.0 + 2.3 = 5.3. As a result of this weighting procedure, the maximal value will be obtained when the 75 dB line passes through the midpoint of the dynamic range.

Central position. The central position of the range is obtained by adding the weighted minimum and maximum distances (from 75 dB) and dividing the sum by two. The minimum and maximum intensities of 55 and 85 dB in the preceding example yield a central position of (-3.0 + 2.3) / 2 = -0.35. The negative sign indicates a central position of the dynamic range below the reference intensity of 75 dB.

Application of the analytical procedure
The application of this structured analysis of phonetograms with algorithms, yielding parameter values, was implemented in a computer program. The application program was written in a fourth-generation signal analysis software package ASYST (MacMillan Software Company), which operates in DOS environment.

Data files proceeding from the computerized registration of phonetogram points consist of a header containing personal data, followed by a clockwise listing of phonetogram points. This file can be used to generate a phonetogram or to serve as the input for the application program. Figure 7 gives a process diagram that summarizes the analytic procedures performed. Processing multiple phonetograms can be done easily by using standardized
Figure 8. The result of the analytic computation is displayed on a monitor. In the upper-right corner the phonetogram is plotted. Underneath the first 30 Fourier Descriptors are given. At the left side the personal data and analyzed parameters are printed.

Figure 9. The phonetogram of a male subject with a mutational voice disorder.

filenames with a numeric part corresponding to a specific subject or patient. This makes possible a highly automated evaluation of a large number of phonetograms when the input for analysis is specified by a starting and ending file number. Figure 8 shows the output displayed on the computer screen. The resulting parameters are displayed on the monitor and stored in an output file. Combining multiple individual output files originating from analyzed phonetograms of subjects without voice complaints, reference parameter values for persons belonging to this group can be established.

Illustration of the application program

To demonstrate the functioning of the program and to illustrate its capability in determining parameters, phonetograms of two male subjects, one without and one with a mutational voice disorder, were processed. The resulting parameters are compared with the mean values and standard deviation of a large group (n=46) of male subjects without voice complaints or voice training, hereafter referred to as the normal reference group (24).

Figure 9 gives the phonetogram of the male subject with a mutational voice disorder, while the phonetogram of the "normal" male subject is used for illustrating the proposed method (see Figures 1 to 8). As far as function is concerned the most salient abnormal aspect of the phonetogram displayed in
Table 1. Parameter values for a "normal" and a pathological (mutational voice disorder) phonetogram with reference values (mean and Standard Deviation [SD]) of a group of 46 male subjects without vocal complaints or vocal training.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>Pathologic</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD_1</td>
<td>.22</td>
<td>.33</td>
<td>.14</td>
<td>.089</td>
</tr>
<tr>
<td>FD_2</td>
<td>.63</td>
<td>.84</td>
<td>.57</td>
<td>.089</td>
</tr>
<tr>
<td>FD_3</td>
<td>.13</td>
<td>.36</td>
<td>.23</td>
<td>.119</td>
</tr>
<tr>
<td>FD_4</td>
<td>.34</td>
<td>.32</td>
<td>.25</td>
<td>.107</td>
</tr>
<tr>
<td>FD_5</td>
<td>.29</td>
<td>.07</td>
<td>.20</td>
<td>.129</td>
</tr>
<tr>
<td>FD_6</td>
<td>.26</td>
<td>.17</td>
<td>.24</td>
<td>.101</td>
</tr>
<tr>
<td>FD_7</td>
<td>.09</td>
<td>.06</td>
<td>.16</td>
<td>.081</td>
</tr>
<tr>
<td>FD_8</td>
<td>.05</td>
<td>.10</td>
<td>.18</td>
<td>.094</td>
</tr>
<tr>
<td>FD_9</td>
<td>.03</td>
<td>.10</td>
<td>.14</td>
<td>.075</td>
</tr>
<tr>
<td>Contour Regularity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope (dB/st)</td>
<td>1.01</td>
<td>.64</td>
<td>.91</td>
<td>.208</td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enclosed area</td>
<td>.24</td>
<td>.11</td>
<td>.25</td>
<td>.052</td>
</tr>
<tr>
<td>Frequency range (# octaves)</td>
<td>3.2</td>
<td>2.8</td>
<td>3.2</td>
<td>.32</td>
</tr>
<tr>
<td>&quot;Speaking Range&quot; Dynamics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WDR_{mff-3 st}</td>
<td>5.3</td>
<td>0.8</td>
<td>3.4</td>
<td>1.84</td>
</tr>
<tr>
<td>WDR_{mff}</td>
<td>5.4</td>
<td>0.5</td>
<td>5.2</td>
<td>1.10</td>
</tr>
<tr>
<td>WDR_{mff+1/2 oct}</td>
<td>5.5</td>
<td>0.7</td>
<td>6.0</td>
<td>.41</td>
</tr>
<tr>
<td>WDR_{mff+1 oct}</td>
<td>5.9</td>
<td>4.2</td>
<td>6.0</td>
<td>.40</td>
</tr>
<tr>
<td>Position_{mff-3 st}</td>
<td>-0.4</td>
<td>-2.5</td>
<td>-1.6</td>
<td>.91</td>
</tr>
<tr>
<td>Position_{mff}</td>
<td>-0.3</td>
<td>-2.8</td>
<td>-0.7</td>
<td>.54</td>
</tr>
<tr>
<td>Position_{mff+1/2 oct}</td>
<td>-0.1</td>
<td>-2.6</td>
<td>-0.2</td>
<td>.21</td>
</tr>
<tr>
<td>Position_{mff+1 oct}</td>
<td>0.3</td>
<td>-0.9</td>
<td>0.1</td>
<td>.19</td>
</tr>
</tbody>
</table>

Figure 9 is the restricted dynamic range at lower frequencies. To indicate specific deviations from "normal" male phonetograms, both phonetograms are analyzed with the mff standardized at 123 Hz (B2 in American notation).

Table 1 gives the analyzed parameter values of the "normal" and pathologic phonetograms, as well as the mean values and standard deviations (SD) of the normal reference group. The "normal" phonetogram yields parameter values all within a range of 2 SDs from the mean value, which is commonly accepted as defining a normal range. The subject with a mutational voice disorder, however, produced a phonetogram that yields deviant parameter values. Regarding the shape, the first and second FD show values above the range of 2 SDs. The enclosed area is small compared to the reference norm. Summarizing the parameters of the "speaking range" dynamics, the weighted
ranges are, except for the lowest sampled frequency, all significantly small and the central positions of these ranges are well below the reference intensity, which means that at all sampled frequencies phonations are only possible with soft intensities. In short, the phonetogram of the subject with vocal problems is abnormal with respect to shape, enclosed area and "speaking range" dynamics, and has parameter values that might conceivably be representative for a mutational voice disorder.

DISCUSSION

The power and robustness of the proposed parameters largely depend on a standardized registration of phonetogram points. Directions and instructions were formulated by Schutte & Seidner (6). However, because the shape parameters are dependent on the number of points in a phonetogram, we strongly advise a consistent choice of points at which the frequency range is sampled. Following Schutte & Seidner, four frequencies per octave are recommended. When a recommended step on either end of the range is beyond the phonatory possibilities of the tested subject, an increment or reduction by a tone or semitone will provide the extremes.

Fourier Descriptors can be used to describe quantitatively a shape. Applying the analytic method proposed by Zahn and Roskies (26) to phonetograms results in an order of Fourier Descriptors with varying amplitudes (see Figure 2c). By processing a large number of phonetograms and averaging amplitudes a "mean" phonetogram can be obtained. Further research is needed to establish specific relationships between one or more Fourier Descriptors.

The shape is influenced by the dB(A) weighting network used to register the sound level (18). The increasing attenuation of frequencies below 500 Hz reduces the SPL levels of the loud and especially the soft contours at the low frequencies. Where phonetogram analysis is used as a method for comparing phonetograms and for observing and detecting changes in phonetograms under the influence of therapy or training, its power as a clinical instrument is not compromised by a standardized use of a dB(A) weighting network. However, this weighting network does limit the scientific use of phonetograms in research on voice function.

Registering phonetograms we accept reproducible phonations at a given frequency with a minimum phonation time of one second. This duration insures a stabilized sound intensity production and correct measurement. A smaller minimum phonation time, as compared with the three seconds recommended by Coleman, Mabis, & Hinson (11) and Schultz-Coulon & Asche (3), reveals the physiological extremes of the voice and makes the procedure more practicable.
Chapter 7

in clinical practice with patients suffering from laryngeal diseases associated with short phonation times.

The description of phonetograms with explicit parameters offers the possibility of determining deviations, from normal values, for pathological entities. If these deviations show specific disease-related characteristics, each type of disease with effects upon phonation and phonatory potentialities could be represented by a set of parameter statistics. Knowledge about a special combination of parameter statistics can be used to build an expert system (28). In an expert system specific knowledge on observed behavioral patterns is formalized in a computer program, giving the users a stronger basis for making decisions. A suggested expert system based on knowledge of constellations of parameter values of phonetograms, specific for groups of subjects or patients, could be a very useful tool for speech pathologists, therapists and ENT-physicians to help in diagnosing diseases or in supporting a diagnosis.

Before this can be realized, a large number of phonetograms per pathologic entity have to be analyzed in order to build the knowledge base. A computerized phonometer with analysis techniques can make a major contribution toward an optimal clinical employment of phonetograms.

References