Acoustic voice analysis by means of the hoarseness diagram

Abstract

The hoarseness diagram (Michaelis, Fröhlich, & Strube, 1998a) has been proposed as a new approach to describe different acoustic properties of voices. To test its performance in the analysis of pathologically disturbed and normal voices five requirements are suggested that should be met by any acoustic voice analysis protocol to be used in voice research and clinical practice. The hoarseness diagram is then tested with regard to these requirements. Individual voices are found to show a satisfactory localization in the diagram. Aspects of stationarity are discussed by four case studies. The different cases illustrate that changes in the acoustic analysis results are observed if the voice generation conditions change while results are stationary if phonation conditions do not change. Different pathological voice groups defined on grounds of the specific phonation mechanism are found to map to specific regions of the hoarseness diagram with differences between group locations being significant. All results can be interpreted without exceptions if the two hoarseness diagram coordinates are taken to reflect the vibrational irregularity of the voice generation mechanisms on the one side and the degree of closure of the vibrating structures on the other side. The hoarseness diagram and its underlying algorithms are thus shown to constitute a useful approach to acoustic voice analysis in research and clinical practice. The tests themselves demonstrate several application possibilities including the quantitative monitoring of individual voices.

Acoustic voice analysis by means of the hoarseness diagram

Introduction

In the quantitative assessment of voice characteristics the use of acoustic features has received an increasing amount of attention in the past years. Several advantages are conceivable in the acoustic analysis of pathological voices. Most prominently, the reproducibility of results and the independence of the experimenter allow an objective characterization of a voice. Acoustic analysis is non-invasive and does not require semi-invasive tools like endoscopes. Therefore subjects may be tested easily for different voice tasks such as sustained phonation of different vowels or running speech. Such an analysis of different voice tasks is important if the every-day performance and range of the voice is to be assessed. Acoustic analysis methods offer these possibilities whereas the assessment of the vocal product may be difficult to realize for different speech tasks with other diagnostic methods such as laryngo-stroboscopy or high-speed imaging of the vocal fold vibration. More pragmatic advantages concern the low costs of the analysis both in terms of equipment and personnel.

Problems in acoustic analyses may arise from contaminations of the acoustic signal. While background noise can generally be controlled as well as a possible Doppler effect (Schoentgen, Bensaid, Bucella, & Ciocea, 1998), the specific vocal tract resonances (Michaelis et al., 1998b; Rammage, Peppard, & Bless, 1992) or the heartbeat cycle (Orlikoff & Baken, 1989) may have an effect on the values of acoustic measures. However, the main drawback seems to be the still unresolved discussion concerning the general interpretation of analysis results in terms of perceptual or physiological correlates (Dejonckere, 1995; Dejonckere & Lebacq, 1996; de Krom, 1995; Eskenazi, Childers, & Hicks, 1990; Feijoo & Hernández, 1990; Hammarberg, Fritzell, Gauffin, Sundberg, & Wedin, 1980; Hammarberg, Fritzell, & Schiratzki, 1981; Hillenbrand & Houde, 1996; Kreiman & Gerratt, 1994; Murry, Singh, & Sargent, 1977; Rammage, Peppard, & Bless, 1992; Södersten & Lindestad, 1990; Zwirner, Michaelis, Fröhlich, Strube, & Kruse, 1998).

Some general correspondences have nevertheless emerged in the interpretation of the acoustic analysis results in relation to the underlying voice generation process. First, irregularities that have been detected in the radiated voice signal by appropriate acoustic features may be attributed to irregularities

of the vocal fold vibration (Dejonckere, 1995). Well-known measures of this category are jitter and shimmer. Second, incomplete glottal closure during the closed phase of the glottal cycle causes turbulent airflow and consequently an increased amount of additive noise in the radiated signal (Dejonckere & Lebacq, 1996).

Acoustic analysis results that are based on the measurement of those two signal characteristics (irregularity and additive noise content) may therefore be interpretable in terms of the underlying voice production process. We use the term *phonation mechanism* to describe the main characteristics of the voice production process (i.e., whether the vocal folds or supra-glottal structures serve to generate the voice, or whether glottal closure can be realized or not). The *phonation condition* further specifies the individual conditions of voice generation(e.g., the local vibration amplitude that – in the case of a tumor – is affected by size and location of the tumor). Different phonation mechanisms can be described in a qualitative way so that voices can be grouped categorically by phonation mechanism. On the other hand, phonation conditions for individual subjects vary on a continuous scale. If groups based on specific phonation mechanism are to be compared, the qualitative difference in voice generation *between* the groups should exceed the differences *within* a group due to the variation of individual phonation conditions. But even then, the choice of appropriate features and algorithms remains crucial so that the usefulness of acoustic features has even been questioned altogether (Bielamowicz, Kreiman, Gerratt, Dauer, & Berke, 1996; Rabinov, Kreiman, Gerratt, & Bielamowicz, 1995).

Still another problem arises in the clinical application of voice analysis systems. It concerns the actual performance of an acoustic analysis system in that it may not analyze voices that are “too disturbed” (e.g., Kay Elemetrics MDVP: “signal not voiced”). Such “dropouts” are particularly irritating in the assessment of voice changes for individual subjects who possess such highly disturbed voices, for example during a voice therapy.

Motivated by the above observations, a list of requirements may be compiled that should be met by an analysis system that is to be used in clinical practice or voice research:

requirement 1 (analyzability): all voices can be analyzed,

requirement 2 (localization of results for individuals): results show a small scattering during one recording session relative to the possible range of values (small intra-subject variability),

requirement 3 (stationarity of results for individuals): (a) results do not change much if phonation conditions remain similar, (b) results change a lot if phonation conditions change,

requirement 4 (localization of results for voice groups): similar phonation mechanisms result in similar values (small inter-subject variability),

requirement 5 (interpretability): results are interpretable in terms of phonation conditions without contradictions.

The somewhat vague formulations should allow to test the requirements at different levels of desired accuracy. While strictly speaking some of the requirements refer to intrinsic properties of the tested data, these data properties should still be reflected by the results of the acoustic analysis.

The initially stated list of potential advantages of acoustic voice analysis in the every-day phoniatric and logopedic routine has sparked a rapid development of measures and commercially available systems in the past years. A relatively recent development in the field of acoustic voice analysis is the *hoarseness diagram* (Fröhlich, Michaelis, & Kruse, 1998a,b; Fröhlich, Michaelis, Strube, & Kruse, 1997; Kruse, Michaelis, Zwirner, & Bender, 1997; Michaelis, Fröhlich, & Strube, 1998a). This diagram is based on four acoustic measures, three of which assess different aspects of signal irregularity and the fourth the additive noise content. The particular measures have been determined on the basis of extensive studies on the informational gain and inter-dependence of various acoustic features (Michaelis et al., 1998a).

In this article the hoarseness diagram provides the data to test the analysis procedure in the light of the above list of requirements. Requirement 1 is not examined explicitly since the acoustic features underlying the hoarseness diagram have been shown in previous studies to meet this requirement (Fröhlich et al., 1997). Requirement 2 is tested by analyzing the scattering of analysis results for individual recording sessions. The examination of changes in the analysis results of individual subjects on a long-term time scale allows an evaluation with regard to requirement 3. Requirement 4 is tested by analyzing various voice groups with defined phonation mechanisms and looking into the localization and statistical differentiation of the group distributions. The most important issue with regard to the potential applicability is given by requirement 5 which influences all other points and is therefore discussed where appropriate.

Methods

Data

A voice corpus of $N = 425$ recordings of subjects with different vocal pathologies of varying severity provided the data for the acoustic analyses (see Table 1). The recordings were taken as part of the clinical routine at the Department of Phoniatrics and Pedaudiology at Göttingen University, Germany. Additionally, 93 recordings of normal voices (no reported history of voice problems, range 11 to 61 years, mean 27 years) as well as 60 recordings of whispered vowels simulating total aphonia were used as reference groups. The speech samples were recorded in a sound-treated room using a head-mounted microphone (beyer dynamics HEM 191), pre-amplifier (AXR Mic/Dat 2), and DAT recorder (Pioneer D-07, sampling frequency 48kHz). During each recording session the subject was instructed to phonate the vowel sequence [ɛ: a: e: i: o: u: ɛ:] first at comfortable pitch and intensity, then at a lower pitch, then at a higher pitch. Next, a standard text ("Nordwind und Sonne", duration approximately 2 minutes) was read (not used for this study) after which the vowel sequence was repeated at comfortable pitch. The subject was instructed to sustain each vowel for three to five seconds, if possible, without connecting consecutive vowels. The pitches of the different vowel series were not further specified or controlled, since subjects with highly disturbed voices were often unable to phonate at different pitches. Only recordings where each of the 28 vowel samples was sustained for at least one second entered the voice corpus. The only data pre-processing that was performed manually was the exclusion of voice onset and offset. The remaining "stationary" part of the signal was analyzed automatically and in a completely unsupervised way.

(insert Table 1 about here)

Acoustic analysis: the hoarseness diagram

The hoarseness diagram allows a quantitative two-dimensional description and graphical representation of voice characteristics on the base of four acoustic measures (Fröhlich et al., 1997, 1998a,b; Kruse et al., 1997; Michaelis et al., 1998a). These four measures were found to yield a low-dimensional description of an originally 21-dimensional acoustic data space with the least loss of information. Principal components analysis revealed that in this four-dimensional space the data were

distributed in approximately a two-dimensional plane. The two principal components describing this plane provided the basis for the two axes of the hoarseness diagram (Michaelis et al., 1998a).

In the diagram, three measures (jitter, shimmer, mean period correlation) contribute to the horizontal axis labeled the *irregularity component* (IC) (Michaelis et al., 1998a, equation 7). All three are based on a segmentation of the acoustic signal into a sequence of glottal cycles. This unsupervised segmentation is performed by the waveform matching method (Milenkovic, 1987; Titze & Liang, 1992). The algorithm yields highest IC coordinates for a signal generated by a random number generator (Michaelis et al., 1998a). By following this algorithmic approach for segmentation any signal can be analyzed and no *ad hoc* threshold for "analyzability" needs to be introduced.

The fourth acoustic measure is the glottal to noise excitation ratio GNE (Michaelis, Gramss, & Strube, 1997). It indicates to what extent the voice excitation is due to a pulse train or due to noise. In tests using synthetic signals the GNE was shown to be sensitive to additive noise but – unlike the normalized noise energy NNE (Kasuya, Ogawa, Kikuchi, & Ebihara, 1986) or cepstral harmonics to noise ratio CHNR (de Krom, 1993) that are often used to assess additive noise in speech signals – to be independent of jitter and shimmer (Michaelis et al., 1997). The GNE enters the hoarseness diagram by the vertical coordinate labeled the *noise component* (NC) (Michaelis et al., 1998a, equation 8). While it is important to keep in mind that all measures assess characteristics of the radiated acoustic signal rather than direct properties of the voice generation process, they can – within certain limitations (Michaelis et al., 1998b) – generally be viewed as correlates of the underlying production mechanisms (Dejonckere, 1995; Dejonckere & Lebacq, 1996).

The voice recordings were analyzed in 500ms frames applying a shift of 250ms. For each complete frame (i.e., generally excluding the last frame of a vowel utterance) the hoarseness diagram coordinates were calculated. The resulting coordinate distribution of a complete recording session was characterized by the mean ($= \mu_{\mathrm{IC}}, \mu_{\mathrm{NC}}$) and standard deviation ($= \sigma_{\mathrm{IC}}, \sigma_{\mathrm{NC}}$) for each coordinate. For visualization an (outlined) ellipse was plotted in the hoarseness diagram with the means defining the center coordinate and the standard deviations defining the half-axes.

For some subjects the voice corpus contained several recordings at different dates. In the assessment of the individual scattering (requirement 2) multiple recordings of the same subject were used for the

pathological group if recordings were at least one week apart. For the normal group, multiple recordings of the same subject at different times were not allowed. This proceeding was motivated by the observation that pathological voices often undergo substantial changes during wound healing or voice therapy. In such cases recordings of the same vocally disturbed subject at different times may be regarded as samples of different pathological voice conditions. In this way the number of data points was increased while the bias that was potentially introduced by the multiple analysis of some voices was kept small.

For the investigations concerning requirement 4 (localization of results for voice groups) only one recording of each subject was used. This choice was based on the phoniatric examination without any reference to the results of the acoustic analysis or to perceptual ratings.

Localization of the analysis results for individuals (requirement 2)

The standard deviations calculated for each recording session were used to characterize the localization of the analysis results for individuals. The ratio of the average standard deviation to the largest distance between ellipse center coordinates – calculated from all recordings available – was chosen as numerical indicator v of the average variability of one recording session. It was calculated for each axis IC and NC ($N = 425$):

$$v = \frac{\frac{1}{N}\sum_{i=1}^{N}\sigma_i}{\max\limits_{\substack{j=1,\dots N\\ k=1,\dots N}}\{|\mu_j - \mu_k|\}} \quad . \tag{1}$$

Stationarity of the analysis results for individuals (requirement 3)

The hoarseness diagram offers the possibility to display analysis results of different recording sessions in the same diagram. Individual changes in the acoustics of a voice can thus be assessed in a quantitative manner. For clinical purposes, this objective description of changes in voice characteristics allows for example to monitor the progress of a voice rehabilitation program. The monitoring of individual voices is illustrated in four case studies.

Subject bw (male, 55 years) had a second glottal tumor resected prior to the first recording which followed an earlier partial bilateral cordectomy. During the following half year he underwent a

functional voice therapy to stabilize the phonation mechanism. Stages of his improvement in voice generation are listed in Table 2. The second subject bs (female, 63 years) suffered from a bilateral Reinke's edema that was resected after the first recording (Table 2, middle). The third subject kw (male, 47 years) had a sub-total resection of a laryngeal tumor (T3) of the right endolarynx and applied an ary-epiglottic phonation mechanism (Table 2, bottom). Subject fm (male, 29 years) possessed a normal voice. The subjects were chosen to illustrate the assessment of changes in acoustic properties for mildly and severely disturbed voices (subjects bs and bw, respectively) and of unchanging phonation conditions for pathological and normal voices (subjects kw and fm, respectively).

(insert Table 2 about here)

In order to assess the normal variability in the analysis results five normal subjects were tested for differences in group distributions (see next section) on a long-term time scale. The first group comprised recordings of the five subjects at the earliest date available in the data base. The second group comprised the most recent recordings of the same subjects. Intervals between first-group and second-group recordings ranged from 7-20 months.

Localization of the analysis results for voice groups (requirement 4)

One possibility to determine whether similar voice generation conditions result in comparable analysis results is to relate results for subjects with well-defined phonation mechanisms to each other. Given the clinical motivation of the analysis method, the first and most obvious grouping of the subjects is by medical diagnosis. However, many diagnoses do not imply specific phonation conditions and thus neither well-defined phonation mechanisms. In these cases a further selection has to be performed in a second step on grounds of the phonoscopic [1] examination.

Groups were tested in two main categories. In the first category different phonation mechanisms encountered after cancer surgery and subsequent voice rehabilitation were compared ("cancer groups"). In the second category two types of laryngeal paralyses were compared ("paralysis groups").

The groups in both categories comprised subjects with pronounced deviation from normal laryngeal conditions. This seemed important for two reasons: a) it allowed to test the performance of the analysis methods for highly disturbed voices where other systems often fail to produce interpretable results, and

b) in these cases groups with homogeneous phonation mechanisms could be defined for the data available.

The normal voice group and a group of simulated aphonic voices (realized by recordings of whispered vowels) marked the most extreme voice generation conditions. They supplied the references necessary for the interpretation and thus for the *a posteriori* validation of the analysis results. The acoustic analysis results for a given group were displayed as (filled) ellipse. Its location and size was defined by the distribution of the ellipse centers of the individual recordings underlying each group: for each coordinate the group ellipse center was defined as the mean of the centers of the individual recordings, the half-axis as the corresponding standard deviation. [2]

Cancer groups

Patients who have suffered from laryngeal carcinoma show post-surgically various types of phonation mechanisms (Kruse, 1998). The particular mechanism depends primarily on the size of the resected tumor but can be also influenced by voice therapy.

Patients with small carcinoma may be able to continue using the vocal folds for voice generation after partial cordectomy and wound healing. Two different situations are observed. In the first case the operated vocal fold still vibrates or vibration can be restored by functional voice rehabilitation (Kruse, 1998). This situation is termed “glottic phonation” (group gp). The other case occurs when tissue properties of the cancerous fold have changed to such a degree that the fold has become stiff. For voices showing this condition the term “pseudo-glottic phonation” is suggested (group pgp). The two different classes of vibration patterns are distinguished by phonoscopic examinations. For pgp subjects the glottal closure is also observed to be not very tight.

After a complete cordectomy or if a major part of the vocal fold had to be resected the ventricular folds may serve as voice source (group vp). For vp subjects the vibration of the ventricular folds is usually highly irregular after the operation while closure is poor. The regularity of the vibration can be improved considerably by voice rehabilitation. Subsequently also the closure of the ventricular folds improves.

If even the ventricular folds cannot be used to generate the voice, ary-epiglottic phonation may still

be possible (group aep). This phonation type involves both epiglottic tissue and part of the arytenoid cartilage or the ary-epiglottic fold to generate the voice. The irregular vibrations observable during phonoscopy can be improved just slightly by voice rehabilitation. Correspondingly, the closure is also relatively poor during the phonation of sustained vowels.

The groups gp, pgp, vp, and aep used for the analyses comprised subjects showing exactly one of the phonation mechanisms just described. If more than one recording existed of a subject, the one after completing voice rehabilitation was used.

Paralysis groups

The different nerves of the larynx may be traumatized independently or in combination. We distinguish two types of nerve damage as described by Kirchner (1977) for subjects with laryngeal paralyses and immobility of the vocal folds.

Most commonly the recurrent nerve is paralyzed. In general, this type of paralysis leads to a paramedian position of the paralyzed fold. In the laryngoscopic view during breathing the paralyzed fold is positioned in almost the normal phonation position. A relatively stable glottal vibration is usually possible due to the entrainment of the paralyzed fold with the healthy one.

Less often a lesion of the vagus nerve between the turnoff of the superior laryngeal nerve and the turnoff of the recurrent nerve is observed. Subjects with this type of paralysis generally show a more lateral respiration position of the paralyzed fold in the laryngoscopic examination. In the phonoscopic examination, however, two distinct conditions can be observed: for most subjects glottal closure is poor during phonation. On the other hand, some subjects consistently realize glottal closure that is almost complete but not very tight (short closed phase).

In order to define homogeneous phonation mechanism groups on the base of the medical diagnosis, a two-step procedure was applied. First, a pre-grouping was performed based on etiology. The diagnosis relied particularly on the laryngoscopic examination assessing the respiratory condition. Only subjects with unilateral paralysis entered the two groups rpa (recurrent nerve paralysis) and vpa (vagus nerve paralysis). If several recordings of a subject existed, the one showing the clearest indication of the paralysis before logopedic treatment (usually the subject's first recording) was chosen on grounds of the

medical examinations.

In the second step, the groups were re-examined giving particular attention to the phonoscopic image sequences. On these grounds the vagus paralysis group was split up into the two groups "vagus paralysis with glottal closure" (group vpa$^+$) and "vagus paralysis without glottal closure" (group vpa$^-$). Since group vpa$^+$ consisted of just two subjects, only group vpa$^-$ was further analyzed.

Statistical significance

The statistical significance of the differences between the distributions underlying the group ellipses was tested by the two-dimensional Kolmogorov-Smirnov test (Press, Flannery, Teukolsky, & Vetterling, 1988). It allowed to test the two-dimensional difference in the plane defined by the hoarseness diagram. In order to test the significance between distribution differences in one dimension (i.e., for the IC or NC) the Wilcoxon two-sample test was applied (Press et al., 1988). The Bonferroni-Holm correction (Holm, 1979) was applied in order to adjust for multiple comparisons ($N = 15, 6, 252$ comparisons in the analysis of the cancer, paralysis, and vowel groups (see Appendix), respectively). Student's t-test for unequal variances (Press et al., 1988) was applied to test the significance of the one-dimensional differences between the centers of the group distributions for both coordinates. This test was applied to assess differences between the normal and the pathological group (req. 2) and changes for individual voices (req. 3). Throughout the paper significances are interpreted at the 5% error level ($p < 0.05$).

Results

Requirement 2: Localization of the results for individuals

Table 3 shows values characterizing the distributions of the ellipse centers and standard deviations for the pathological and the normal voice groups. Since subjects had to sustain each vowel for at least one second, each ellipse was based on at least 84 data points (average 305). Additional analysis results concerning particular speech tasks such as the phonation of different vowels or of voice tasks at different pitches are supplied in the Appendix.

(insert Table 3 about here)

The distributions of the standard deviations were found to be unimodal and relatively symmetrical.

For the pathological voice group the average standard deviation was 1.04 | 0.58 in the IC | NC. For the normal group the average standard deviation was smaller in the IC (0.84) but not in the NC (0.61). Student t-test for unequal variance revealed that the difference in the IC was significant whereas in the NC it was insignificant. The average standard deviation calculated for all recordings of both groups together was 1.02 | 0.59 in the IC | NC.

The v-ratios according to equation (1) for the two axes were calculated to be $v_{\mathrm{IC}} = 0.14$ and $v_{\mathrm{NC}} = 0.17$. These values indicate a satisfactory localization of individual ellipses in relation to the complete range of observed values for both coordinates.

Requirement 3: Stationarity for individuals

For three subjects with pathologically disturbed voices short descriptions of the phonoscopic examinations were supplied in Table 2. The corresponding changes in individual voice acoustics are shown in the top three diagrams of Fig. 1. The bottom diagram shows the acoustic variations of a normal voice on a long-term time scale.

(insert Figure 1 about here)

Subject bw (Table 2, top) possessed post-surgically on 3/13/96 a highly disturbed but not aphonic voice. No oscillations of laryngeal structures could be seen in the stroboscopic examinations. Acoustic analysis results of this date show very high IC and NC values (Fig. 1, top). In the first stages of the subsequent voice rehabilitation vibration of the vocal folds became apparent in the stroboscopic examinations with a phase difference between the left and right fold. On 6/26/96 both folds were vibrating while a pronounced glottal gap was still observed. The acoustic analysis during this time shows a continuous decrease in IC with hardly any changes in NC. After this date the vibration further improved, showing by a decreasing phase difference between the oscillating folds. Glottal closure improved as well but remained incomplete in the anterior part. The acoustic analysis shows a continuous decrease in both coordinate values during this later part of the logopedic rehabilitation.

The voice generation of subject bs (Table 2, middle) changed from irregular vibration and incomplete closure before the resection to a slightly asymmetric vibration of both vocal folds and an almost complete glottal closure after voice rehabilitation. Acoustically, her voice showed a continuous

development during wound healing and logopedic therapy (Fig. 1, second from top).

Subject kw showed the same ary-epiglottic phonation conditions in all phonoscopic examinations during two years (Table 2, bottom). In the acoustic analyses (Fig. 1, third from top), all ellipses are found at more or less the same location (standard deviation of the different ellipse centers 0.09 in both coordinates). The consistency of the location of the ellipses illustrates that changes in the acoustic properties are not necessarily observed for pathological voices. Instead, acoustic analysis results are located in the same region of the hoarseness diagram once a stable phonation mechanism has developed.

The results for subject fm (Fig. 1, bottom) illustrate the acoustic fluctuations of a normal voice during two years. All ellipses are found in more or less the same location (standard deviation of the ellipse centers 0.21 | 0.11 in the IC | NC).

In order to estimate the sensitivity of the hoarseness diagram to changes in voice conditions statistical tests were performed on the acoustic analysis results of the different dates. For each subject the differences between the centers of the ellipses were tested by the Student t-test for unequal variances using the underlying data distributions of the particular recording sessions. For subject bw, all differences were significant, with the only exception for the ellipse pair {5/24/96, 5/31/96}. Since the subject received logopedic voice therapy and the recordings were taken at intervals of roughly two weeks, the small but monotonic changes in the analysis results may be interpreted as reflecting the changes in voice generation mechanism due to the therapy. The high frequency of significantly different ellipse centers demonstrates the high sensitivity of the hoarseness diagram when assessing functional changes in voice generation for individual subjects.

For subject bs all differences between ellipse centers were significant. Her case demonstrates that the hoarseness diagram may be used to assess changes in acoustic voice properties not only for severe voice disturbances but also for subjects with slightly or moderately impaired voices if the regularity of the vibration or the degree of glottal closure is affected.

The acoustic analysis results for subject kw revealed significant differences between ellipse centers for all possible pairs but for {11/19/97, 6/17/98} despite the close proximity of all ellipses. Similarly, for subject fm 146 of the 171 possible ellipse pairs showed significantly different center coordinates. This high frequency of significant differences can be explained by the large amount of data underlying each

ellipse (222 | 366 data points on average for subject kw | fm). For the vocally healthy subject fm the observed variations in size and location indicate the general variability of a voice due to environmental factors (e.g., air humidity, pollen) or health conditions (recording with a sore throat on 1/20/97). Equivalent minute differences during the phonoscopic examination cannot be expected to be measurable without a reliable quantitative image analysis of the vocal fold motion.

The average normal variability that was tested by the group analysis of the five normal subjects on a long-term time scale revealed differences between group centers of 0.01 | 0.06 in the IC | NC. Both differences were insignificant both in the two-dimensional and one-dimensional tests (Kolmogorov-Smirnov test and Wilcoxon test, respectively).

Requirement 4: Localization of analysis results for voice groups

Cancer groups

The acoustic analysis results for the different phonation mechanism groups after cancer resection and subsequent voice rehabilitation are displayed in Fig. 2 together with the reference groups of the normal and aphonic voices. Results of the statistical analyses are shown in Table 4.

(insert Figure 2 and Table 4 about here)

All acoustic analysis results for the different groups – including the reference groups – were found to be significantly different from each other. Differences in vibrational regularity of the voice source and in degree of closure of the vibrating structures are thus reflected by the acoustic analysis results. Furthermore, all cancer patients showed considerable changes in tissue properties due to the carcinoma or scars obtained during the resection, which explains the significant differences from the normal group. The significant differences between the cancer groups and the aphonic reference group reflect the fundamental difference between a voiced phonation source and aphonia. In this regard, the difference between the ary-epiglottic phonation group (aep) and the aphonic voices is particularly interesting. Two of the five subjects of the aep group could not be analyzed by standard commercial software (Kay Elemetrics MDVP: “signal not voiced”), the other three only in small parts of the signal. Yet ary-epiglottic phonation represents a voiced sound generation and in the phonoscopic examination a vibration of tissue is generally observable. Although the group aep is situated close to the aphonic group

in the hoarseness diagram (Euclidean distance of means 1.23) the significant difference in the acoustic analysis results reflects this functional difference in voice generation.

In patho-physiological respect, groups gp and pgp are closest to normal phonation since the vocal folds still serve as voice source. Of the two groups, gp is closer to normal phonation than pgp since in the former case both folds oscillate while in the latter situation only one fold vibrates. Still, the vocal folds serve to generate the voice in both groups. Of the two remaining groups, vp can be considered closer to "normal" than aep for three reasons. For ventricular phonation, oscillation patterns are observed that are comparable to vocal fold vibration in terms of orientation and symmetry. Second, a surface wave can be seen that is similar to the mucosal wave of the vocal folds, and third, less laryngeal tissue has generally been resected.

In the acoustic domain, analysis results may be ordered with increasing Euclidean distance from the origin. This sequence reads as "normal – gp – pgp – vp – aep – aph" and thus reflects the underlying voice generation properties just described. Additionally, for pgp subjects the scarred vocal fold that does not vibrate leads to a poor glottal closure which is reflected by a high NC value for the pgp group ellipse relative to gp. The pattern found in the acoustic analysis results can thus be interpreted in terms of the patho-physiological deviation in laryngeal conditions from the normal voice.

Paralysis groups

The results of the acoustic analyses for the different paralysis groups are shown in Fig. 3 and Table 5. Statistical analysis of the group distributions shows that the difference between the groups rpa and vpa$^-$ is significant as well as the differences between the two paralysis groups and the reference groups. The significant differences from the reference groups were to be expected since a unilateral vocal fold paralysis represents a relatively severe voice disturbance which would account for the significant difference from normal voices. On the other hand, in the first stages of the paralysis a vibration of the affected fold is usually still possible. This would explain the significant difference from the aphonic voices.

(insert Figure 3 and Table 5 about here)

The significant differences in the acoustic analysis results between the two paralysis groups may be

explained – as in the case of the cancer groups – by the different patho-physiologic phonation mechanisms. Poor glottal closure leads to air turbulences that show acoustically in additive noise and correspondingly in elevated NC values. An increased NC value is indeed observed for the voices of group vpa$^-$ where voices do not show glottal closure by definition. The absence of glottal closure is a result of a weaker or missing entrainment between the oscillating healthy vocal fold and the paralyzed one. This reduces the overall regularity of the speech signal since the paralyzed fold is "flapping" in an unsynchronized way which is observed during phonoscopy. The reduced regularity is reflected by the higher IC center coordinates for group vpa$^-$.

Discussion

The hoarseness diagram is a new approach to the acoustic analysis of pathological voices in that it combines several acoustic measures on grounds of theoretical considerations and statistical analyses (Michaelis et al., 1998a). Jitter, shimmer, and mean period correlation belong to the well-known group of irregularity measures, GNE to the measures of additive noise. Although at least jitter and shimmer have been used in many studies, the specific algorithms and the particular combination of measures are crucial both in obtaining the maximum information in the two dimensions of interest (Michaelis et al., 1998a) and in achieving meaningful analysis results.

Requirement 1 states that all voices should be analyzable. Given the motivation for the two dimensions of the hoarseness diagram, two extreme conditions of the analyzed voice signal have to be considered: (a) extreme noise content, (b) extreme irregularity. The GNE was tested for both conditions in Michaelis et al. (1997). The behavior of irregularity measures for condition (b) has been documented in Fröhlich et al. (1997) and Michaelis et al. (1998b), for condition (a) in Hillenbrand (1987). The analysis results of aphonic voices confirm that this most extreme case of voice disturbance also leads to the highest coordinate values observed. The possibility to analyze highly disturbed voices and even aphonic voices on a continuous scale without introducing an *ad-hoc* threshold for unvoicedness is due to the specific design of the algorithms (Fröhlich et al., 1997). The hoarseness diagram therefore compares favorably to most commercially available acoustic analysis systems or other common acoustic measures (Fröhlich, Michaelis, & Strube, 1998c) that do not allow the analysis of highly disturbed voices. With

regard to the list of postulated requirements, requirement 1 is met.

Requirement 2 concerns the localization of analysis results for individuals in relation to the possible value range. This localization is influenced by several factors such as the variation of articulatory constellations, pitch, or intensity for the different tokens of one recording session (some additional analyses concerning these issues are discussed in the Appendix). The average standard deviation for a recording session of a pathological voice was 1.0 | 0.6 in the IC | NC (Table 3). The significantly smaller standard deviation for the normal group in the IC (0.84) corresponds to the expectations that the vibration of the vocal fold is generally more stable for normal voices than for pathologically disturbed voices.

The average standard deviation for the pathological group is much higher than the test-retest variability on a long-term time scale for the two subjects fm and kw with stable phonation conditions (0.2 | 0.1 in the IC | NC). For the different vowels, the maximum distance between the average ellipse centers was 1.3 | 0.85 in the IC | NC (see Appendix). The large average size of the ellipses for individual recordings (1.0 | 0.6 in the IC | NC) and the similar average ellipse sizes for the normal and the pathological group in the NC (0.6 for both groups, difference insignificant) may therefore be attributed primarily to the mixing of different vowel tokens. Consequently, acoustic analysis results should only be compared if the same recording protocol was used. Scherer, Vail, & Guo (1995) suggested at least 15 analyzed vowel tokens for acoustic analyses. For the present study the advantage of obtaining a very low test-retest variation by using all 28 vowel tokens outweighed the increase in individual ellipse size that resulted from the pooling of different vowels. Comparing the average ellipse size to the range of ellipse centers by the v-values of equation (1) (0.14 | 0.17 in the IC | NC), requirement 2 can be considered met even with the observed mean standard deviation for individual recording sessions of 1.0 | 0.6 in the IC | NC.

Requirement 3 concerns aspects of stationarity in the analysis results for individual subjects. The four case studies illustrate the robustness of the hoarseness diagram results and their interpretability according to functional aspects of voice generation. For each of the two subjects with unchanging voice generation conditions (kw, fm) the analysis results were located in close proximity. Acoustic analysis results on a long-term time scale for a group of normal subjects showed only minute differences that were all statistically insignificant. The hoarseness diagram therefore meets requirement 3a. On the other

hand, in the cases of the two subjects bw and bs with changing voice generation conditions the acoustic analysis results reflect these changes. These subjects demonstrate the performance of the hoarseness diagram with regard to requirement 3b.

The high frequency of significantly different ellipse centers found for all four subjects indicates that slight changes in the voice signal between different recording sessions can be captured by the hoarseness diagram. Since small fluctuations in voice generation conditions may lead to small differences that are nevertheless significant, acoustic analysis results should be interpreted with reference to actually observed changes in voice generation conditions (i.e., in reference to the phonoscopic findings). If the voice generation has changed within a certain time for a given subject, significant changes in the acoustic analysis results may be interpreted to reflect these changes as long as irregularity of the vibration and closure are affected. The high frequency of significant differences indicates that in these cases the assessment of changes in voice properties may be possible with a high certainty of measurements.

Requirement 4 was tested by relating acoustic analysis results of subject groups with similar phonation mechanisms to each other. Highly pathological phonation mechanism groups were used since in these cases the patho-physiological differences between groups exceeded the intra-group variability due to individual phonation conditions. The analysis results showed that groups with homogeneous phonation mechanisms were well-localized in the hoarseness diagram. Significant differences of the group locations were observed between different phonation mechanisms, which suggests that requirement 4 is met. Further analyses using quantitative description of voice generation – either by a parametric classification of the phonoscopic images or by image analysis of high-speed phonoscopic recordings – is needed to further substantiate these findings.

The attempt to test the hoarseness diagram results in the light of the initially stated list of requirements was motivated by the possibility of its clinical application. The approaches to test the particular requirements also originated in a clinical setting. This led to certain limitations in the design of the tests. While requirements 1 and 2 could be addressed adequately on grounds of the data available, the tests of requirements 3-5 should be regarded as first steps only. More formal tests supplying further validation will require a quantitative assessment of the voice generation process. These quantitative data could be supplied by a parametric description of the vocal fold vibration. Additionally, data obtained

from inverse filtering by use of a flow mask might be used additionally to test the correlation between DC-flow and NC values. However, ultimately the image analysis of high-speed recordings of the laryngeal vibrations should serve to test requirements 3-5 in quantitative detail. Such additional tests will remain interesting topics for future studies.

Both parametric descriptions or high-speed image analyses of the vocal fold vibration would allow to compile additional groups with homogeneous phonation mechanisms of similar phonation conditions. For moderately disturbed voices subject groups with comparable phonation conditions are hard to define on grounds of qualitative data only. For example, in the case of a benign tumor the specific effect on voice generation strongly depends on size and position of the tumor. By using a quantitative description of the voice generation process not only groups with highly disturbed voices could be examined (such as the cancer or paralysis groups of this study) but also subject groups showing slight and moderate voice impairments.

When analyzing moderately impaired voices with the hoarseness diagram it has to be kept in mind that voice disturbances not primarily affecting the degree of glottal closure or the regularity of vocal fold vibration could not be expected to lead to significant differences from normal voices. Also, a decrease in the NC – corresponding to a more complete glottal closure – may not always be a goal of a voice therapy since many normal voices also exhibit glottal gaps (Södersten et al., 1995). Trying to lower NC coordinate values by increasing the degree of glottal closure might ultimately result in a hyper-functional voice. For voices that are only mildly disturbed a higher dimensional description of acoustic signal characteristics might be more appropriate (Hammarberg et al., 1980; Michaelis et al., 1998a).

Informal observations suggest that within a given category (cancer or paralysis) predictions concerning the underlying phonation mechanisms may be possible on grounds of the acoustic analysis results. Future studies have to confirm this formally. On the other hand, if the general type of voice disturbance is unknown, predictions concerning the voice generation condition are not advisable since voices with very different etiologies might share regions in the two-dimensional acoustic space described by the hoarseness diagram.

Conclusion

The hoarseness diagram was found to show a reliable localization of individual analysis results. Stationarity of the acoustic analysis results was observed if phonation conditions did not change. On the other hand, changes in voice generation conditions corresponded to changes in the acoustic analysis results.

In the analysis of different pathological voice groups with well-defined phonation mechanisms the groups were found to map to certain regions of the hoarseness diagram. The differences between the group distributions were in all cases statistically significant within each class of voice pathology. For some groups the medical diagnosis itself did not suffice to compile groups showing homogeneous voice generation conditions (such as undifferentiated “vagus paralysis” or “vocal fold phonation” groups). In these cases the final classification into groups with homogeneous phonation mechanisms relied particularly on the phonoscopic examinations.

Without exception, all results were interpretable if the irregularity component IC was taken to reflect the vibrational irregularity of the voice generation mechanism (usually the vocal folds) and the noise component NC the degree of closure of the vibrating structures. The hoarseness diagram was thus shown to be a useful acoustic analysis method by the initially postulated requirements of analyzability, localization of the results for individuals and voice groups, stationarity of the results for individuals, and an unequivocal interpretation.

References

Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., & Berke, G. S. (1996). Comparison of voice analysis systems for perturbation measurement. *Journal of Speech and Hearing Research, 39,* 126–134.

de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research, 36*, 224–266.

de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research, 38*, 794–811.

Dejonckere, P. (1995). Principal components in voice pathology. *Voice, 4*, 96–105.

Dejonckere, P. & Lebacq, J. (1996). Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology. *ORL; Journal of Oto-Rhino-Laryngology and its Related Specialities*, *58*(6), 326–332.

Eskenazi, L., Childers, D., & Hicks, D. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research, 33*, 298–306.

Feijoo, S. & Hernández, C. (1990). Short-term stability measures for the evaluation of vocal quality. *Journal of Speech and Hearing Research, 33*, 324–334.

Fröhlich, M., Michaelis, D., & Kruse, E. (1998a). Image sequences as necessary supplement to a pathological voice data base. In G. de Krom (Ed.), *Procedings of VOICEDATA98* (pp. 64–69). Utrecht: Utrecht Institute of Linguistics OTS.

Fröhlich, M., Michaelis, D., & Kruse, E. (1998b). Objektive Beschreibung der Stimmgüte unter Verwendung des Heiserkeits-Diagramms. *HNO, 46*, 684–689.

Fröhlich, M., Michaelis, D., & Strube, H. W. (1998c). Acoustic “breathiness measures” in the description of pathologic voices. In *Proceedings ICASSP 98* (pp. 937–940). Seattle, WA.

Fröhlich, M., Michaelis, D., Strube, H. W., & Kruse, E. (1997). Acoustic voice quality description: Case studies for different regions of the hoarseness diagram. In T. Wittenberg, P. Mergell,

M. Tigges, & U. Eysholdt (Eds.), *Advances in Quantitative Laryngoscopy, 2nd 'Round Table'* (pp. 143–150). Erlangen: Dept. Phoniatrics.

Fröhlich, M., Michaelis, D., Strube, H. W., & Kruse, E. (1998d). Stimmgütebeschreibung mit Hilfe des Heiserkeits-Diagramms: Untersuchung verschiedener pathologischer Gruppen. In M. Gross (Ed.), *Aktuelle phoniatrisch-pädaudiologische Aspekte 1997/98* (pp. 42–48). Heidelberg: Median Verlag.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Oto-Laryngologica*, *90*, 441–451.

Hammarberg, B., Fritzell, B., & Schiratzki, H. (1981). Teflon injection in 16 patients with paralytic dysphonia - perceptual and acoustic evaluations. *Speech Transmission Laboratory – Quaterly Progress and Status Report*, *1*, 38–57.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech and Hearing Research*, *30*, 448–461.

Hillenbrand, J. & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, *39*, 311–321.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Kasuya, H., Ogawa, S., Kikuchi, Y., & Ebihara, S. (1986). An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology. *Speech Communication*, *5*, 171–181.

Kirchner, J. (1977). Intrathoracic injury to the motor nerve supply of the larynx. *Acta Oto-Laryngologica*, *83*, 163–169.

Kreiman, J. & Gerratt, B. R. (1994). The multidimensional nature of pathologic vocal quality. *The Journal of the Acoustical Society of America*, *96*(3), 1291–1302.

Kruse, E. (1998). Therapie der Stimm-, Sprech- und Sprachstörungen. In G. Böhme (Ed.), *Sprach-, Sprech, Schluck- und Stimmstörungen* (pp. 114–130). Stuttgart: Fischer Verlag.

Kruse, E., Michaelis, D., Zwirner, P., & Bender, E. (1997). Stimmfunktionelle Qualitätssicherung in der kurativen Mikrochirurgie der Larynxmalignome auf der Basis der „Laryngealen Doppelventilfunktion". *HNO, 45*, 712–718.

Martin, D., Fitch, J., & Wolfe, V. (1995). Pathologic voice type and the acoustic prediction of severity. *Journal of Speech and Hearing Research, 38*, 765–771.

Michaelis, D., Fröhlich, M., & Strube, H. W. (1998a). Selection and combination of acoustic features for the description of pathologic voices. *The Journal of the Acoustical Society of America, 103*(3), 1628–1639.

Michaelis, D., Fröhlich, M., Strube, H. W., Kruse, E., Story, B., & Titze, I. R. (1998b). Some simulations concerning jitter and shimmer measurement. In T. Lehmann, C. Palm, K. Spitzer, & T. Tolxdorff (Eds.), *Advances in Quantitative Laryngoscopy, Voice and Speech Research. Proceedings of the 3rd International Workshop* (pp. 71–80). Aachen: RWTH University of Technology. ISBN 3-00-002945-1.

Michaelis, D., Gramss, T., & Strube, H. W. (1997). Glottal-to-noise excitation ratio – a new measure for describing pathological voices. *Acustica / acta acustica, 83*, 700–706.

Milenkovic, P. (1987). Least mean square measures of voice perturbation. *Journal of Speech and Hearing Research, 30*(4), 529–538.

Murry, T., Singh, S., & Sargent, M. (1977). Multidimensional classification of abnormal voice qualities. *The Journal of the Acoustical Society of America, 61*(6), 1630–1635.

Orlikoff, R. F. & Baken, R. J. (1989). Fundamental frequency modulation of the human voice by the heartbeat: Preliminary results and possible mechanisms. *The Journal of the Acoustical Society of America, 85*(2), 888–893.

Press, W., Flannery, B., Teukolsky, S., & Vetterling, W. (1988). *Numerical Recipes in C*. Cambridge: Cambridge University Press.

Rabinov, C. R., Kreiman, J., Gerratt, B. R., & Bielamowicz, S. (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research*, *38*, 26–32.

Rammage, L. A., Peppard, R. C., & Bless, D. M. (1992). Aerodynamic, laryngoscopic, and perceptual-acoustic characteristics in dysphonic females with posterior glottal chinks: A retrospective study. *Journal of Voice*, *6*(1), 64–78.

Scherer, R. C., Vail, V. J., & Guo, C. G. (1995). Required number of tokens to determine representative voice perturbation values. *Journal of Speech and Hearing Research*, *38*, 1260–1269.

Schoentgen, J., Bensaid, M., Bucella, F., & Ciocea, S. (1998). Issues in the acoustic evaluation of laryngeal disorders. In G. de Krom (Ed.), *Procedings of VOICEDATA98* (pp. 16–21). Utrecht: Utrecht Institute of Linguistics OTS.

Södersten, M., Hertegård, S., & Hammarberg, B. (1995). Glottal closure, transglottal airflow , and voice quality in healthy middle-aged women. *Journal of Voice*, *9*(2), 182–197.

Södersten, M. & Lindestad, P.-Å. (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech and Hearing Research*, *33*, 601–611.

Titze, I. & Liang, H. (1992). Comparison of F_0 extraction methods for high precision voice perturbation measurements. *National Center for Voice and Speech – Status and Progress Report*, *3*, 97–115.

Zwirner, P., Michaelis, D., Fröhlich, M., Strube, H. W., & Kruse, E. (1998). Korrelationen zwischen perzeptueller Beurteilung von Stimmen nach dem RBH-System und akustischen Parametern. In M. Gross (Ed.), *Aktuelle phoniatrisch-pädaudiologische Aspekte 1997/98* (pp. 63–67). Heidelberg: Median Verlag.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft under Kr 1469/2-2. Parts of the results have been presented at the *Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie 1997* (Fröhlich, Michaelis, Strube, & Kruse, 1998d). We thank Dr. R. Orlikoff, Dr. J. Sundberg, and the anonymous reviewers for their helpful suggestions regarding earlier versions of this manuscript.

Authors' addresses:

Matthias Fröhlich, Dirk Michaelis, Hans Werner Strube

Drittes Physikalisches Institut, Georg-August Universität Göttingen

Bürgerstr. 42-44

D-37073 Göttingen, Germany

[matth | micha | strube]@physik3.gwdg.de

Eberhard Kruse

Dept. Phoniatrics and Pedaudiology, Georg-August Universität Göttingen

Robert-Koch Str. 40

D-37075 Göttingen, Germany

ekruse@med.uni-goettingen.de

Appendix

Throughout this study the acoustic voice analysis of a given subject at a certain date was based on all 28 vowels recorded per session. Since those 28 tokens do not represent strictly homogeneous data, additional analyses on token groups were performed to get a more detailed picture of the analysis results in general. Two general inhomogeneities were looked into: (a) the mixing of articulatory constellations (i.e., different vowels), (b) the mixing of different pitches.

(a) In order to test the influence of the articulatory configuration the different vowels of each recording session were analyzed separately before being averaged for the normal and the pathological group (Fig. 4). Results of the statistical tests are stated for the pathological group in Table 6.

(insert Figure 4 and Table 6 about here)

For both groups a similar pattern is observed. [ɛ:] and [a:] are located at highest IC and lowest NC values, [i:] and [e:] at slightly lower IC but similar NC values, [o:] and [u:] at low IC and high NC values. The differences in standard deviation between the vowel distributions were generally insignificant. The significances of the differences between the individual vowel distributions in Table 6 point to two different effects. In the one-dimensional analyses the rounded vowels [o:] and especially [u:] differ markedly and significantly from the others in the NC. On the other hand, the IC values separate the low vowels [ɛ:], [a:] from the high vowels [i:], [o:], [u:].

The significantly increased NC values of the rounded vowels [o:] and [u:] can be explained by the specific properties of the GNE. Informally it was observed that subjects often tended to articulate these vowels with a very pronounced rounding of the lips. Sometimes this even led to a whistling character of the signal. This demonstrates that the airflow at the lips was high so that a generation of air turbulences at the lip constriction seems reasonable. Acoustically these turbulences would show in additive noise. Together with the strong attenuation of the higher harmonics this would decrease the GNE thereby increasing the NC values.

A possible explanation for the differences in the IC values is harder to conceive. Tests with synthesized speech samples described in Michaelis et al. (1998b) showed a strong influence both of the vowel and the fundamental frequency F_0 on the regularity of the radiated signal. Those tests were performed with a source-filter model without the possibility of feedback on the glottis model. The

observed vowel patterns do not show an F_0 dependency. It might be speculated that the deviation from the source-filter simplification, i.e. the possibility of negative feedback in real voice production, might be responsible for the observed differences in regularity. The higher energy density within the vocal tract for the closed vowels might increase the repercussions on the vocal fold vibration. This would stabilize the vibration and could thus explain the lower IC values.

(b) In order to test the influence of the pitch the four vowel series (abbreviated as normal1, low, high, normal2) were analyzed separately before being averaged for the normal and the pathological group. The differences in standard deviation between the series amounted to maximally 0.08 and were all insignificant. Comparing the distributions of the centers, all differences were significant except for the pairs {normal1,normal2} (both groups), {normal1,low} (both groups), {high,normal2} (pathological group only). In the one-dimensional analysis no significant differences were found in the NC. Ordered with increasing IC values the same sequence “high – normal2 – normal1 – low” resulted in both groups. Voice properties with regard to the regularity of the signal were thus affected by the instruction to phonate at different pitches. This is in accordance with other studies describing effects of the fundamental frequency on irregularity measures and/or the perceived roughness (de Krom, 1995; Martin, Fitch, & Wolfe, 1995). It should be noted that the pitch dependence was observed even though the realization of different pitches was not controlled. Nevertheless, the mean fundamental frequencies show that on average a variation of the pitch was realized (for the pathological | normal group: 170 | 171Hz (first series in normal pitch), 156 | 146Hz (low pitch series), 220 | 225Hz (high pitch series), 177 | 174Hz (second series in normal pitch)).

The mean difference in center coordinates between the two normal pitched series was 0.1 in both coordinates for both groups. This value is comparable to the variation of analysis results for individual subjects with stable phonation conditions on a long-term time scale (2 years) as described in the main body of the text under requirement 3. It indicates a low test-retest variation on the time scale defined by the duration of the recording session (less than 10 minutes).

Footnotes

[1]: The term “phonoscopy” is used to describe the endoscopic examination of the larynx during phonation with the possibility to resolve individual (pseudo-)cycles. This observation of the dynamical behavior may be realized by laryngo-stroboscopy in the case of periodic vibration. For irregular vibration, it should be performed preferably by high-speed imaging or kymography. Although for the examinations performed in this study the term “phonoscopy” always refers to laryngo-stroboscopy, it is used to draw attention to the fact that the possibility to resolve the dynamical behavior is crucial for this particular examination. The term “laryngoscopy”, on the other hand, refers to the general endoscopic examination of the larynx without the possibility to resolve the dynamical behavior during phonation.

[2]: The group ellipse is based on the distribution of the individual ellipse centers only. Neither the sizes of the individual ellipses for the different recording sessions nor the corresponding underlying shapes of the data distribution enter the group ellipse.

Table 1: Occurrence of voice disorders of the 425 recordings used for analysis (258 different subjects, mean age 48 years, range 12 to 80 years). The categories are based on etiology. Some subjects carry multiple diagnoses.

number of occurrences	diagnosis/description
18	functional dysphonia
12	pre-operative status before micro-surgery (glottal carcinoma)
199	post-operative status after partial laryngectomy
46	vocal fold paralysis
14	vocal fold immobility
9	mutational disorders
87	benign tumors
27	psychosomatic dysphonia
42	others (maximum 4 per diagnosis)

Table 2: Description of phonoscopic examinations.

Subject	Date	Description
bw	3/13/96	distinct fibrinous layer in anterior commissure (wound healing); compensatory activity of ventricular folds; no vibration of vocal folds observable during stroboscopy
	6/26/96	phase difference between vocal folds (vibration of left fold better than of right fold); poor closure
	9/4/96	vibration of both vocal folds with a slight phase difference between left and right fold; incomplete closure in anterior third
bs	9/12/96	pre-operative: incomplete closure, irregular vibration; Reinke's edema: large (3) on right fold, intermediate (2) on left fold
	10/21/96	post-operative: incomplete closure; reduced vibrational ability of right fold (especially at high pitches), phase difference between the folds; small edema on left fold, minimal fibrinous cover on right fold
	12/13/96	pre-rehabilitation: slight phase difference between folds (less than for 10/21/96); slightly reduced vibrational ability of right fold; improved closure compared to 10/21/96 but still incomplete
	2/18/97	post-rehabilitation: almost symmetrical vibration of left and right fold; very slight glottal gap in the anterior part but otherwise complete closure
kw	4/17/96	irregular vibration between left ary-epiglottic fold and petiolus; mucus
	5/21/97	irregular vibration between left ary-epiglottic fold and petiolus
	6/17/98	irregular vibration between left ary-epiglottic fold and petiolus

Table 3: Average means and standard deviations of the analysis results of individual recording sessions. Means, standard deviations (s.d.) and ranges of the ellipse centers and half-axes are stated for both coordinates (n: number of recordings).

group	n	irregularity component				noise component			
		average	s.d.	min	max	average	s.d.	min	max
ellipse centers									
pathological	425	4.79	1.78	1.76	9.18	2.29	0.93	0.54	3.93
normal	37	2.98	0.47	2.34	4.00	1.22	0.54	0.46	2.69
ellipse half-axes									
pathological	425	1.04	0.35	0.34	2.42	0.58	0.23	0.18	1.14
normal	37	0.84	0.21	0.49	1.49	0.61	0.23	0.17	1.09

Table 4: Cancer group: Two-dimensional tests of significance (Kolmogorov-Smirnov test) of the differences between the groups glottic phonation (gp), pseudo-glottic phonation (pgp), ventricular phonation (vp), ary-epiglottic phonation (aep), normal voices, and aphonic voices. The numbers state the Euclidean distances between the group centers. All pairwise differences between group distributions are significant ($p < 0.05$, Bonferroni-Holm correction applied for $N = 15$).

	gp	pgp	vp	aep	aphonia ($n = 60$)
normal ($n = 93$)	1.16	3.50	4.10	5.69	6.56
gp ($n = 15$)		2.34	3.00	4.59	5.45
pgp ($n = 9$)			1.23	2.55	3.33
vp ($n = 10$)				1.59	2.46
aep ($n = 6$)					0.88

Table 5: Paralysis groups: Two-dimensional tests of significance (Kolmogorov-Smirnov test) of the differences between the recurrent paralysis group (rpa), the vagus paralysis group without glottal closure (vpa$^-$), the normal group, and the aphonic group. The numbers state the Euclidean distances between the group centers. All pairwise differences between group distributions are significant ($p < 0.05$, Bonferroni-Holm correction applied for $N = 6$).

	rpa	vpa$^-$	aphonia ($n = 60$)
normal ($n = 93$)	1.02	5.60	6.56
rpa ($n = 24$)		4.63	5.56
vpa$^-$ ($n = 5$)			1.05

Table 6: Two-dimensional and one-dimensional Euclidean distances between the centers of the ellipses of the different vowels for the pathological group. In the one-dimensional analyses a negative sign signifies $\alpha < \beta$. Statistical tests of the group differences were performed by the Kolmogorov-Smirnov test (two dimensions) and the Wilcoxon test (one dimension). Insignificant differences ($p < 0.05$) are indicated by a dagger (†). Bonferroni-Holm correction was applied for the complete set of tests on the distributions of the centers and standard deviations for the normal and pathological voice group ($N = 252$).

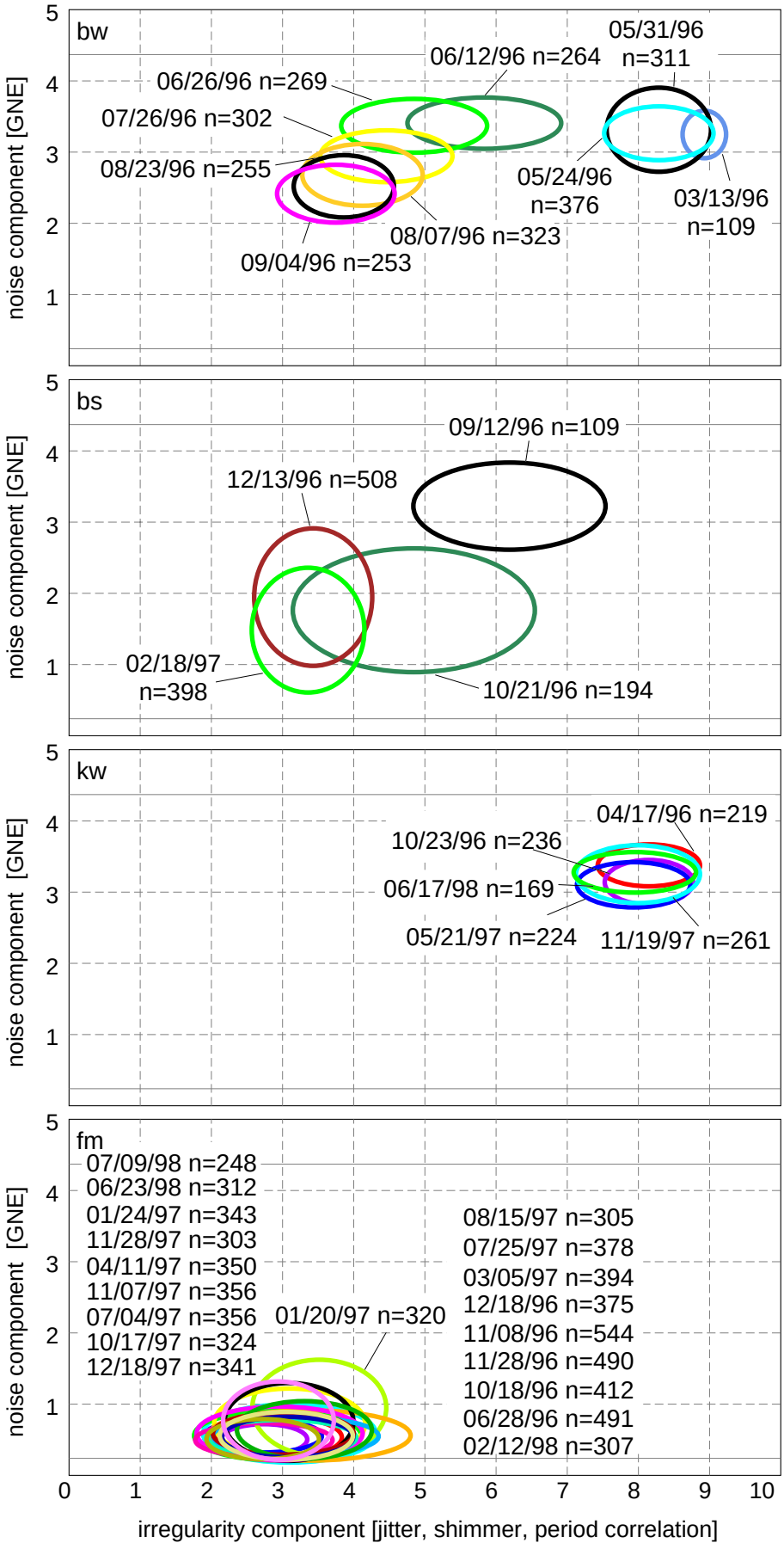
($\alpha \downarrow, \beta \rightarrow$)	[a:]	[e:]	[i:]	[o:]	[u:]	[ɛ:]
two dimensions of the hoarseness diagram						
[ɛ:]	$0.28^\dagger$	$0.36^\dagger$	0.71	0.83	1.33	$0.03^\dagger$
[a:]		0.63	0.97	1.04	1.49	$0.27^\dagger$
[e:]			$0.36^\dagger$	0.56	1.08	$0.37^\dagger$
[i:]				$0.30^\dagger$	0.80	0.71
[o:]					0.53	0.82
[u:]						1.32
one dimension: irregularity component						
[ɛ:]	$-0.26^\dagger$	$0.36^\dagger$	0.70	0.72	1.03	$-0.01^\dagger$
[a:]		0.62	0.97	0.99	1.29	$0.26^\dagger$
[e:]			$0.34^\dagger$	$0.37^\dagger$	0.67	$-0.37^\dagger$
[i:]				$0.02^\dagger$	$0.33^\dagger$	-0.71
[o:]					$0.30^\dagger$	-0.73
[u:]						-1.04
one dimension: noise component						
[ɛ:]	$-0.10^\dagger$	$0.01^\dagger$	$-0.11^\dagger$	-0.41	-0.84	$-0.03^\dagger$
[a:]		$0.11^\dagger$	$-0.01^\dagger$	-0.31	-0.74	$0.07^\dagger$
[e:]			$-0.12^\dagger$	-0.42	-0.85	$-0.04^\dagger$
[i:]				-0.30	-0.73	$0.08^\dagger$
[o:]					-0.43	0.38
[u:]						0.81

Figure 1. Analysis results for four subjects. Subject bw (top) was treated for laryngeal cancer and applies glottic phonation with vibration of the operated vocal fold. Subject bs (second from top) had a resection of a Reinke's edema. Subject kw (third from top) uses ary-epiglottic phonation after sub-total resection of the right endolarynx. Subject fm (bottom) possesses a normal voice. The recording dates and the number n of analyzed frames underlying each ellipse are stated. Descriptions of the laryngoscopic examination corresponding to particular dates are supplied in Table 2.

Figure 2. Group ellipses in the hoarseness diagram for different phonation mechanisms (glottic phonation, pseudo-glottic phonation, ventricular phonation, ary-epiglottic phonation) of subjects after the resection of laryngeal carcinoma. The normal and aphonic voice groups are shown as references.

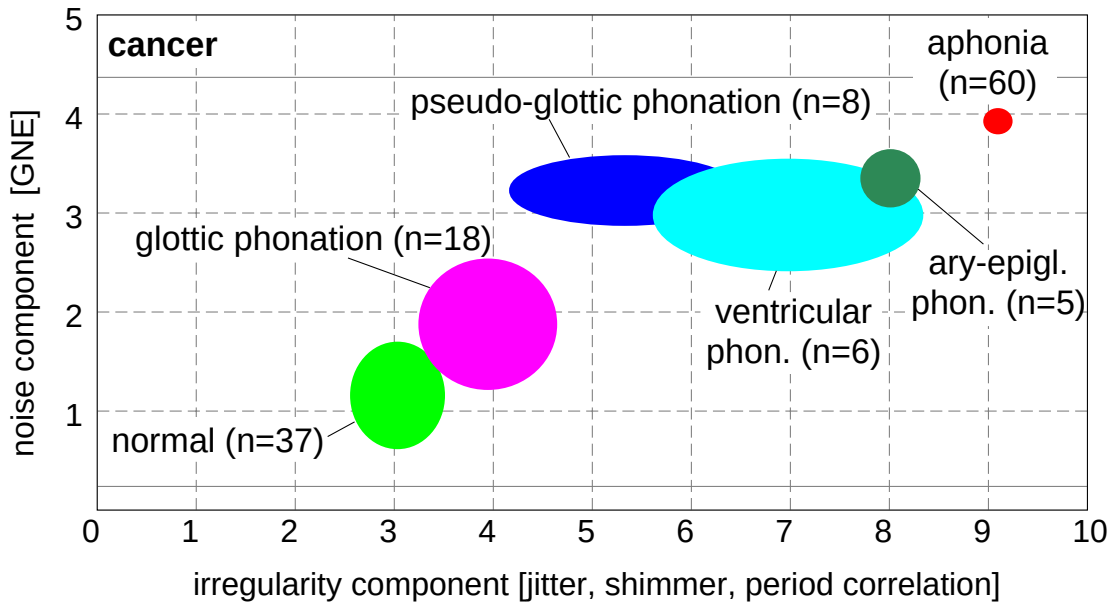
Figure 3. Group ellipses in the hoarseness diagram for subjects diagnosed with recurrent paralysis and vagus paralysis without glottal closure. The normal and the aphonic voice groups are shown as references.

Figure 4. Vowel ellipses. The ellipses of the different vowel tokens were averaged for the normal and the pathological voice groups.


bw
06/12/96 n=264
05/31/96
n=311
06/26/96 n=269
07/26/96 n=302
08/23/96 n=255
05/24/96
n=376
03/13/96
n=109
08/07/96 n=323
09/04/96 n=253
bs
09/12/96 n=109
12/13/96 n=508
02/18/97
n=398
10/21/96 n=194
kw
04/17/96 n=219
10/23/96 n=236
06/17/98 n=169
05/21/97 n=224
11/19/97 n=261
fm
07/09/98 n=248
06/23/98 n=312
01/24/97 n=343
11/28/97 n=303
04/11/97 n=350
11/07/97 n=356
07/04/97 n=356
10/17/97 n=324
12/18/97 n=341
01/20/97 n=320
08/15/97 n=305
07/25/97 n=378
03/05/97 n=394
12/18/96 n=375
11/08/96 n=544
11/28/96 n=490
10/18/96 n=412
06/28/96 n=491
02/12/98 n=307
noise component [GNE]
irregularity component [jitter, shimmer, period correlation]
0
1
2
3
4
5
6
7
8
9
10

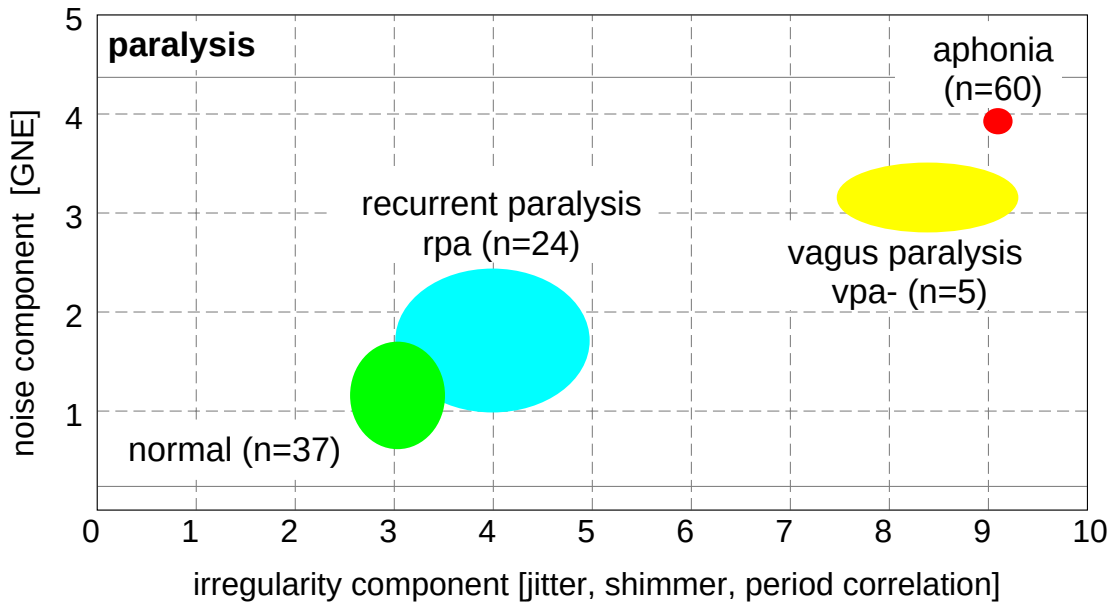
Fröhlich, Figure 1

ACOUSTIC ANALYSIS WITH THE HOARSENESS DIAGRAM


cancer
aphonia
(n=60)
pseudo-glottic phonation (n=8)
glottic phonation (n=18)
ary-epigl.
phon. (n=5)
ventricular
phon. (n=6)
normal (n=37)
noise component [GNE]
irregularity component [jitter, shimmer, period correlation]
0
1
2
3
4
5
6
7
8
9
10

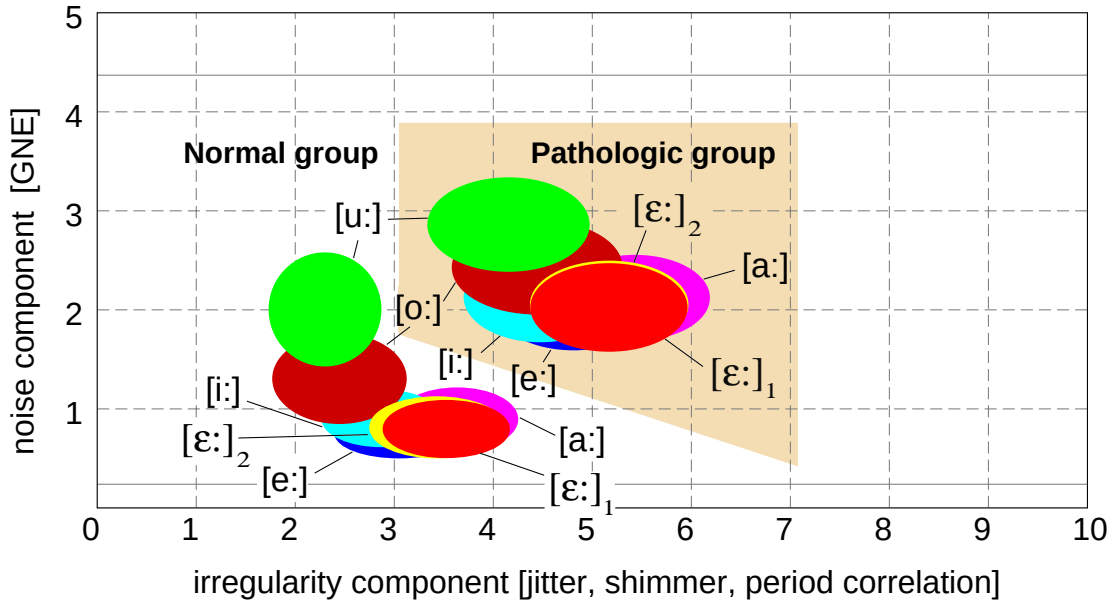
Fröhlich, Figure 2

ACOUSTIC ANALYSIS WITH THE HOARSENESS DIAGRAM


paralysis
aphonia
(n=60)
recurrent paralysis
rpa (n=24)
vagus paralysis
vpa- (n=5)
normal (n=37)
noise component [GNE]
irregularity component [jitter, shimmer, period correlation]
0
1
2
3
4
5
6
7
8
9
10

Fröhlich, Figure 3

ACOUSTIC ANALYSIS WITH THE HOARSENESS DIAGRAM


Normal group
Pathologic group
[u:]
[ε:]2
[a:]
[o:]
[i:]
[e:]
[ε:]1
[i:]
[ε:]2
[a:]
[e:]
[ε:]1
noise component [GNE]
irregularity component [jitter, shimmer, period correlation]
0
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5

Fröhlich, Figure 4

ACOUSTIC ANALYSIS WITH THE HOARSENESS DIAGRAM