Selection and combination of acoustic features for the description of pathologic voices

Dirk Michaelis, Matthias Fröhlich, and Hans Werner Strube

Drittes Physikalisches Institut, Universität Göttingen, Bürgerstr. 42-44, D-37073 Göttingen, Germany

The glottal to noise excitation ratio (GNE) is an acoustic measure designed to assess the amount of noise in a pulse train generated by the oscillation of the vocal folds. So far its properties have only been studied for synthesized signals where it was found to be independent of variations of fundamental frequency (jitter) and amplitude (shimmer). On the other hand, other features designed for the same purpose like NNE (normalized noise energy) or CHNR (cepstrum based harmonics to noise ratio) did not show this independence. This advantage of the GNE over NNE and CHNR as well as its general applicability in voice quality assessment is now tested for real speech using a large group of pathologic voices (n=447). A set of four acoustic features is extracted from a total of twenty-two mostly well-known acoustic voice quality measures by correlation analysis, mutual information analysis, and principal components analysis. Three of these measures are chosen to assess primarily different aspects of signal aperiodicity while the fourth one indicates the noise content of the signal. All analysis methods lead to the same feature set that consists of a measure of period correlation, jitter, shimmer, and GNE. The two-dimensional projection of this set named "hoarseness diagram" allows a graphical illustration of voice quality that can be easily interpreted.

PACS numbers: 43.70.D, 43.70.G, 43.70.J, 43.72.A

I Introduction

The use of acoustic features in the description of pathological voice quality has been tested in various contexts and with a variety of goals. Some of its attractiveness stems from the idea that they might supply a way to quantitatively assess voice characteristics that are otherwise difficult to measure (e.g., Kreiman and Gerratt, 1996). Studies on pathological voices have correlated acoustic features with perceptual qualities (Murry et al., 1977; Hammarberg et al., 1980, 1981; Fritzell et al., 1983a; Askenfelt and Hammarberg, 1986; Hirano et al., 1986, 1988; Eskenazi et al., 1990; Rammage et al., 1992; Kreiman et al., 1992; Kreiman and Gerratt, 1994; Dejonckere, 1995; de Krom, 1995; Bielamowicz et al., 1996; Hillenbrand and Houde, 1996) or, to a lesser extent, with physiologic conditions at the glottis (Hirano et al., 1986; Rammage et al., 1992). However, the results are often ambiguous and sometimes even contradictory, so the choice of the appropriate acoustic measures as well as their interpretation are still unsolved problems.

Acoustic features may be grouped according to the signal characteristics they are supposed to measure. Although there are many different possible categories, the terms "aperiodicity features" and "noise features" can be considered as two important labels. "Aperiodicity features" have been used to describe perceptual roughness (Hirano et al., 1988; Hillenbrand, 1988; Arends et al., 1990; Dejonckere, 1995) or the periodicity of glottal vibration (Dejonckere, 1995). "Noise features" have been found by some researchers to be indicators of breathiness (Hammarberg et al., 1981; Klatt and Klatt, 1990; Hillenbrand et al., 1994; Dejonckere, 1995; Hillenbrand and Houde, 1996), although these findings are controversial. Also, "noise features" have been related to certain physiological conditions during the phonatory cycle such as a glottal gap (Södersten and Lindestad, 1990) or softer closure (Hillenbrand et al., 1994).

"Aperiodicity features" are designed to capture the various forms of periodicity disturbances in the acoustic signal (Klingholz, 1987). One special kind of aperiodicity is related to changes in the waveform shape between glottal cycles. This effect can be measured by the mean correlation coefficient calculated for all pairs of successive cycles. Two other kinds result from variations of the fundamental frequency (F_0) and of the cycle-to-cycle peak amplitude or energy. These aperiodicities are described by the "classical" features jitter and shimmer, respectively. In the long history of the description of signal aperiodicity many different definitions of jitter and shimmer measures have evolved (for overviews see e.g., Kasuya et al., 1986a; Pinto and Titze, 1990; Bielamowicz et al., 1996). In spite of their widespread application in voice quality assessment, recent findings of Schoentgen and deGuchteneere (1997) may lead to the necessity to rethink the general concept underlying these two acoustic features.

To the "noise feature" group belong the features designed to measure the relative noise component in a speech signal. Prominent members of this group are the harmonics to noise ratio (HNR; Yumoto et al., 1982) and the normalized noise energy (NNE; Kasuya et al., 1986a) that have been studied in various contexts (Hirano et al., 1988; Eskenazi et al., 1990; Childers and Lee, 1991; Kreiman et al., 1993; Dejonckere, 1995; Bielamowicz et al., 1996; Qi and Hillman, 1997). A recently developed feature belonging to this group is the glottal to noise excitation ratio (GNE; Michaelis and Strube, 1995).

One of the main problems with grouping acoustic features according to their principal use is that most of them are sensitive to several acoustic properties. This mutual dependence may be one of the reasons for the difficulty in interpreting seemingly contradictory results found in the literature. With regard to the two groups mentioned above, it is obvious that

waveform correlation, jitter, and shimmer (especially amplitude shimmer) will be noise sensitive to some extent depending on the particular algorithm. On the other hand, NNE and HNR were found to be sensitive to jitter and shimmer for synthetic signals (Michaelis et al., 1997a).

The question now arises, which are the features to supply the *best independent assessment* of irregularity and additive noise in pathologic voices. Only if these two voice properties can be measured independently should conclusions be drawn about their relevance with regard to different perceptual qualities like roughness or breathiness or to the underlying physiologic conditions at the glottis.

In this paper we address this issue, with particular emphasis on whether GNE is the "noise feature" supplying more additional information about a pathological voice than NNE and HNR. We will proceed as follows: first, correlations are calculated for a total of 22 voice quality measures (13 of aperiodicity, 9 of noise). Second, using a technique from information theory, a mutual information analysis is performed to find the best jitter and shimmer measures of our list. Once these features have been determined, mutual information analysis is used again to find the best additional "noise measure". Third, the underlying dimensionality of the thus obtained four-feature sub-space is determined by principal components analysis (PCA). After it is found to be two-dimensional, PCA of the complete 22-dimensional feature space is used to confirm the former results. Finally, a two-dimensional graphical representation named "hoarseness diagram" is derived that allows an easy interpretation of the acoustic features.

II Methods

A Speech material

The German vowel [ɛ:] was recorded for 447 different speakers (male and female) between the ages of 10 and 80 (mean: 48). They showed a variety of organic and functional voice disorders (see Table I). 88 normal voices (persons with no history of voice problems) aged 18 to 90 (mean: 47) were recorded as reference group. The vowel was sustained by the subject at comfortable pitch and loudness for several seconds. It was digitally recorded in a soundproof room using the Kay Computer Speech Lab (CSL 4300) at a sampling frequency of 50kHz. 1s of the middle part of the signal was used for acoustic analysis which was performed in a completely automated and unsupervised way. No segments were rejected since perturbation measures were extrapolated to highly disturbed voices as will be described in the following section. This data set is the one used in the experiments unless stated otherwise.

In order to increase the significance of the mutual information analyses, a second data set was generated. Here the same 447 recordings of the first set were used plus additional recordings of the same patients taken at other times. For the resulting 1099 recordings, the acoustic features described in the following were calculated on 500ms frames with a shift of 250ms. In this way a total of n=13414 analyzed segments (i.e. data points) were obtained. The Wilcoxon test showed no significant differences in the feature distributions between the regular pathological group (n=447) and the large pathological group (n=13414).

Table I: List of functional and physiological disorders.

number of occurrences	diagnosis/description
56	vocal fold paresis
22	vocal fold fixation
34	post-operative status after partial laryngectomy
21	pre-operative status before micro surgery (glottal carcinoma)
39	polyps
16	nodules
19	laryngeal granuloma
28	cysts
12	mutational dysphonias
11	papillomas
24	Reinke's Oedema
47	hypo functional dysphonia
45	status/dysphonia after micro surgery (benign tumors)
11	laryngitis
62	others (less than 5 patients per diagnose)

B Acoustic features

1 Jitter, shimmer, and period correlation

The energy sequence $E(\nu)$ for a periodic signal is given by the sum of the squared sample values of the ν^{th} glottal cycle ($\nu = 0, .., N-1$ (last complete cycle)). For the present study the glottal cycle length $P(\nu)$ of the ν^{th} period is determined by the waveform matching algorithm (Milenkovic, 1987; Titze and Liang, 1992). The time range to be tested by this algorithm is set to 0.5 and 1.5 times the time lag (restricted to 2.5ms$\leq \delta_t \leq$15ms) of the maximum of the auto correlation function calculated for the current frame.

As jitter and shimmer measures the perturbation factor (PF; Hollien et al., 1975) and the perturbation quotient (PQ; Koike, 1971) were chosen. They come as part of commercial analysis systems such as the CSL system of Kay Elemetrics and allow a good comparability of the results to other studies because of their wide-spread application. For a sequence $u(\nu)$ the following definitions were used according to Kasuya et al. (1993):

$$\text{PF} = \frac{100\%}{N-1} \sum_{\nu=1}^{N-1} \left| \frac{u(\nu) - u(\nu-1)}{u(\nu)} \right| \tag{1}$$

$$\text{PQ} = \frac{100\%}{N-K} \sum_{\nu=\frac{K-1}{2}}^{N-\frac{K-1}{2}-1} \left| \frac{u(\nu) - \frac{1}{K}\sum_{k=-\frac{K-1}{2}}^{\frac{K-1}{2}} u(\nu+k)}{\frac{1}{K}\sum_{k=-\frac{K-1}{2}}^{\frac{K-1}{2}} u(\nu+k)} \right| \tag{2}$$

If $u(\nu)$ is chosen as the period length sequence $P(\nu)$, the *period perturbation factor* (PPF) and the *period perturbation quotient* (PPQ) define different jitter measures. In the following, PPF will be abbreviated as j2. For PPQ, different choices of K (3,5,7,11,15) lead to the jitter measures abbreviated as j3, j5, j7, j11, j15 (see Table II). Analogously, shimmer is measured by the *energy perturbation factor* (EPF) (abbreviated as s2) or the *energy perturbation quotient* (EPQ) if $u(\nu)$ is chosen as $E(\nu)$. Again, different values for K result in the measures s3, s5, s7, s11, s15 (see Table II). Since the shimmer measures are based on the energy sequence, they are expected to be considerably less susceptible to noise than the amplitude shimmer

Table II: List of the acoustic features. For each feature the monotonic transformation used to obtain an approximately normal distribution is stated. Mean and standard deviation (s.d.) are calculated separately for the normal and the pathological group. Symbols and abbreviations (see also text): MWC - mean waveform matching coefficient; PPF (EPF) - period (energy) perturbation factor, PPQ (EPQ) - period (energy) perturbation quotient, GNE - Glottal to noise excitation ratio, NNE - normalized noise energy, CHNR - cepstrum based harmonics to noise ratio, n - number of recordings, log - base 10 logarithm.

feature	symbol x	description	unit	transformation y = f(x)	mean(y) normal ($n = 88$)	s.d. (y) normal ($n = 88$)	mean (y) pathologic ($n = 447$)	s.d. (y) pathologic ($n = 447$)
MWC	MWC			log (1-x)	-2.021	0.335	-1.614	0.574
jitter	j2	PPF	%	log x	-0.492	0.234	-0.089	0.611
	j3	PPQ K=3	%	log x	-0.792	0.246	-0.374	0.645
	j5	PPQ K=5	%	log x	-0.734	0.203	-0.323	0.626
	j7	PPQ K=7	%	log x	-0.673	0.211	-0.276	0.610
	j11	PPQ K=11	%	log x	-0.588	0.206	-0.210	0.584
	j15	PPQ K=15	%	log x	-0.522	0.199	-0.161	0.563
shimmer	s2	EPF	%	log x	0.572	0.212	0.848	0.421
	s3	EPQ K=3	%	log x	0.268	0.224	0.550	0.424
	s5	EPQ K=5	%	log x	0.347	0.199	0.617	0.407
	s7	EPQ K=7	%	log x	0.403	0.204	0.662	0.398
	s11	EPQ K=11	%	log x	0.476	0.203	0.717	0.384
	s15	EPQ K=15	%	log x	0.531	0.204	0.757	0.368
GNE	gne1	1000Hz bandwidth		log (1-x)	-1.612	0.291	-1.062	0.515
				x	0.969	0.022	0.834	0.189
	gne2	2000Hz bandwidth		log (1-x)	-1.360	0.331	-0.870	0.485
				x	0.940	0.056	0.768	0.222
	gne3	3000Hz bandwidth		log (1-x)	-1.120	0.345	-0.690	0.428
				x	0.892	0.106	0.695	0.242
NNE	nne1	60-5000Hz	dB	x	-19.425	3.634	-16.025	5.853
	nne2	60-2000Hz	dB	x	-22.831	3.606	-19.492	6.830
	nne3	1000-5000Hz	dB	x	-11.715	3.734	-7.441	4.448
CHNR	chnr1	60-5000Hz	dB	x	25.169	3.649	20.088	7.001
	chnr2	60-2000Hz	dB	x	29.157	3.833	23.877	8.261
	chnr3	1000-5000Hz	dB	x	17.345	4.123	11.609	5.397

often used in acoustic voice analyses.

Titze and Liang (1992) came to the conclusion that jitter and shimmer cannot be determined accurately for highly disturbed voices. Indeed, for increasingly aperiodic signals those features more and more loose their meaning as indicators of deviations from *periodicity*. However, the waveform matching algorithm does not depend a-priori on the degree of signal periodicity and therefore can be applied to any kind of voice signal. For whispered voices (i.e. totally aperiodic signals) the algorithm is found to position the period markers in a seemingly random manner (Fröhlich et al., 1997). Therefore PF and PQ show high values for these voices. This behavior is found to be consistent for any intermediate deviations from periodicity and therefore allows an interpretable quantitative classification of any voice (Michaelis et al., 1997b; Fröhlich et al., 1997, 1998a,b).

The mean of all correlation coefficients evaluated for every pair of consecutive periods is used as the acoustic measure termed *mean waveform matching coefficient* (MWC). It indicates the overall similarity between the cycles of the time signal. The algorithmic evaluation of the MWC is relatively robust compared to the jitter and shimmer calculation. Its upper limit of 1 is reached for signals with identical period shapes (i.e. strictly periodic signals). MWC decreases with increasing differences in length or shape between consecutive periods, i.e. with short-term variations of the time signal.

2 NNE, CHNR, and GNE

Normalized noise energy (NNE) (Kasuya et al., 1986b) and *cepstrum based harmonics to noise ratio* (CHNR) (de Krom, 1993) are designed to measure the relative noise content of a signal. Both NNE and CHNR were found to be sensitive to jitter and shimmer for synthetic signals (Michaelis et al., 1997a). A recently developed feature designed to measure the additive noise in a speech signal is the *glottal to noise excitation ratio* (GNE) (see Appendix).

The use of three different bandwidths leads to different GNE measures (see Table II): gne1 is calculated using a bandwidth of 1kHz, gne2 using 2kHz, and gne3 using 3kHz. Similarly, several different realizations of NNE and CHNR were tested: nne1 and chnr1 are calculated for the frequency range 60 to 5000Hz, nne2 and chnr2 for 60 to 2000Hz, and nne3 and chnr3 for 1000 to 5000Hz.

C Rank correlation analysis

Correlation analysis was applied to the set of 22 acoustic measures stated in Table II in order to determine interdependencies between different features. The statistic applied was the Spearman rank-order correlation coefficient (Press et al., 1989) which is independent of the shape of the underlying data distribution. Pairwise correlations were calculated for all combinations of the acoustic measures ($n = 0.5 \cdot 22 \cdot 21 = 231$). For further interpretations the significance of the difference between correlation coefficients was calculated by

Fisher z-transformation of the correlation coefficients and testing the difference for being normal distributed (van denBrink and Koele, 1987).

The significance levels for all correlations and differences of correlations were calculated applying the Bonferroni-Holm correction (Holm, 1979) which was performed separately for the normal and pathological voice group. The number of tests used for the correction was 369 for the pathological and 261 for the normal group. These totals result from the 231 correlations for each group, 30 ($= 2 \cdot 0.5 \cdot 6 \cdot 5$) differences of all intra-jitter/intra-shimmer correlations (i.,e. correlations of all combinations of the 6 jitter (shimmer) measures) and the {j3,s15} correlation for each group, and – for the pathological group only – 108 ($= 3 \cdot 0.5 \cdot 9 \cdot 8$) differences between all the correlations of the "noise measures" with j3, s15, and MWC (the specific choice of the measures for these tests will become apparent in the specific sections). Throughout this paper results will be interpreted at a significance level of $p \leq 0.05$.

D Mutual information analysis

1 Generalization of the mutual information

The calculation of the mutual information (Fraser and Swinney, 1986) between different acoustic features is based on their probability distributions. Its derivation will be explained in detail for the simple case of a two-dimensional data distribution. Throughout this section, for a given measure the frequency of occurrence within a certain value range will be interpreted as the probability of the measure to be found in this interval.

For the two acoustic features x_1 and x_2 the data are binned to M intervals. These intervals are uniformly distributed along each feature axis and cover the whole corresponding value ranges. Let $n_i(x_1)$ be the number of x_1-values to be found in bin i (analogously $n_i(x_2)$ for feature x_2) while $n_{ij}(x_1, x_2)$ is the number of x_1-values found in bin i and x_2-values found in bin j simultaneously. Using this binned description of the data distribution for x_1 and x_2, the estimated probability to find feature x_1 or x_2 in bin i is given by $p_i(x_\eta) = n_i(x_\eta)/M; i = 1, .., M; \eta = 1, 2$. The estimated joint probability that x_1 will be found in bin i and x_2 in bin j is given by $p_{ij}(x_1, x_2) = n_{ij}(x_1, x_2)/M^2; i, j = 1, .., M$. Using these expressions the one-dimensional and the two-dimensional entropy are given by:

$$H(x_1) = -\sum_{i=1}^{M} p_i(x_1)\mathrm{ld}(p_i(x_1)) \quad (3)$$

$$H(x_1, x_2) = -\sum_{i,j=1}^{M} p_{ij}(x_1, x_2)\mathrm{ld}(p_{ij}(x_1, x_2)) \quad (4)$$

with ld denoting the base 2 logarithm.

With definitions (3), (4) the mutual information between two features is expressed by $I_2 = I(x_1, x_2) = H(x_1) + H(x_2) - H(x_1, x_2)$. I_2 can be interpreted as the average number of bits that are predictable of x_2 if x_1 is known, and vice versa. It can be generalized easily to μ dimensions:

$$I_\mu = I(x_1, .., x_\mu) = \sum_{i=1}^{\mu} H(x_i) - H(x_1, ..., x_\mu). \quad (5)$$

The number of bins M should be as high as possible in order to yield a good approximation of the μ-dimensional probability distribution. However, it is limited by the data available since the number of μ-dimensional hyper-bins necessary to cover the whole μ-dimensional feature space is M^μ. This means that for e.g. $\mu = 4$ and $M = 8$ (which represent values actually used in the analysis) at least 4096 data points are necessary to guarantee statistically the occurrence of at least one point per hyper-bin.

2 Normalized increase of information

If a measure is added to a given set of measures, generally not all bits of the new measure can be predicted by the old ones. The remaining, unpredictable part can be regarded as additional information about the data that is described exclusively by the added measure. However, the interpretation of this informational gain is up to the experimenter as the resulting numerical value does not tell whether the gain is due to random (and therefore unpredictable) noise or whether it indeed indicates new, meaningful properties of the data. In the present study, all acoustic features are known to describe different aspects of voice quality in a meaningful way. Therefore the quantitative value of the informational gain can be used to determine the (in this sense) optimal four-feature set starting from a three-feature set.

The maximum information B (in units "bit") for one feature is $B = \mathrm{ld}(M)$. Using Eq. (3), (5) ($\mu = 3, 4$) we define the *normalized additional information* that results from adding x_4 to the set $\{x_1, x_2, x_3\}$ similar to the *marginal redundancy* of Kumar and Mullik (1996) as

$$\Delta I_R = \frac{B - (I_4 - I_3)}{B}. \quad (6)$$

The mutual information has been shown to be independent of monotonic coordinate transformation (Fraser and Swinney, 1986). Therefore replacing the data values by their ranks as performed in this study is permissible. If the feature values are in this way uniformly distributed, the one-dimensional entropy is maximum: $H(x_i) = B; i = 1, ..., \mu$. Now a value of $\Delta I_R = 0$ indicates that the fourth feature does not add new information (e.g., if it is chosen as one of the features x_1, x_2, or x_3 already present). On the other hand, $\Delta I_R = 1$ indicates the maximum increase of information possible. [1]

E Principal components analysis

The underlying dimensionality of a data distribution as well as its most significant components can be determined by principal components analysis (PCA). The technique applied

[1] Since B is a function of M, the definition of ΔI_R in a strict sense is given by $\Delta I_R = \lim_{M\to\infty}(B - (I_4 - I_3))/\mathrm{ld}(M)$. Due to the finite amount of data this equation cannot be applied, but care has to be taken to use enough data points for Eq. (6).

in this study is the singular value decomposition (SVD; Press et al., 1989). SVD projects correlated measures onto the same principal axis of the data distribution. If the variances in some principal directions are negligible, the data space can be described in the new, lower-dimensional coordinate system defined by the principal vectors.

One advantage of this linear method is that the dimensionality of the data distribution can be estimated simply by counting the number of principal components that account for most of the data variance. The description in a low-dimensional coordinate system often facilitates the interpretation of the data.

All 13 "aperiodicity measures" show asymmetric distributions. Jitter and shimmer are found with a much higher probability at small values, while for MWC most values are close to 1. However, Gaussian data distributions are preferable for principal components analysis. On the other hand, monotonic transformations do not influence the other two analysis techniques (rank correlation analysis and mutual information analysis). Therefore the transformed measures using the different monotonic functions stated in Table II were used for all the analyses.

In order to apply the SVD the data has to be normalized. This is done by first subtracting the mean of the group (pathological/ normal as stated in Table II), then dividing the result by the corresponding standard deviation.

III Results and discussion

A Acoustic analysis of normal and pathologic voices

Table II states means and standard deviations (s.d.) for each measure for the normal and the pathological group. The means of the jitter measures are of similar magnitude. The same is found for the shimmer measures. Compared to jitter, shimmer means are about 10 times higher which shows in differences of approximately 1 for the logarithmic values stated in the table.

The mean GNE increases with decreasing bandwidth which is in accordance with results on synthesized signals (Michaelis et al., 1997a). NNE and CHNR show an estimated noise energy that is about 8dB higher for the frequency range 1 to 5kHz than for 60 to 5000Hz due to the relative attenuation of the harmonics at higher frequencies with regards to the noise level.

The differences between the means of the pathological and the normal group are significant for all measures according to the Wilcoxon two-sample test. Nevertheless, the s.d. for all measures are higher in the pathological group than in the normal group. At the same time the means are less than one s.d. of the pathological group apart for any given measure. This overlap signifies that a discrimination of voice quality based on just one measure cannot be considered complete or unique.

B Rank correlation analysis

1 Correlations between jitter and shimmer

Rank correlations for all combinations of jitter and shimmer measures are significant. For the pathological voice group the correlation between the different jitter features is very high (0.84 to 0.99). For the different shimmer features these correlations are even exceeded (0.90 to 0.99). The correlation between jitter and shimmer measures is still relatively high, but clearly lower (0.74 to 0.87). The lowest correlation (0.74) is found between j3 and s15.

For the normal voices the correlations are generally lower than for the pathologic group but still significant. They range from 0.66 to 0.98 for jitter, 0.73 to 0.98 for shimmer, and 0.46 to 0.76 for the correlation between jitter and shimmer measures. The lowest correlation (0.46) is found again for {j3,s15}.

Since for both groups j3 and s15 are correlated the least, this measure combination supplies the most information on the signal (given this particular set of measures). Therefore, the differences of all intra-jitter (i.e., all combinations of jitter measures) and intra-shimmer were tested against the correlation of this set. [2] For the pathological group, all intra-jitter/intra-shimmer correlations are found to be significantly higher than the {j3,s15} correlation (0.74). For the normal group, the correlations are significantly higher than the {j3,s15} correlation (0.46) with the exception of {j3,j11}, {j3,j15}, and {s3,s15} (correlations 0.75, 0.66, 0.73 respectively).

These findings suggest that for pathological voices the "locality" of the jitter or shimmer measure (i.e. the number of averaged cycles) is of secondary importance. The significant correlations between jitter and shimmer indicate that both measures assess similar voice characteristics. This was to be expected on theoretical grounds since they represent different aspects of signal aperiodicity (Klingholz, 1987) that may be caused by the same phenomenon (e.g. an asymmetry of the vocal folds due to pathological tissue changes) and since jitter automatically introduces shimmer to a signal (Hillenbrand, 1987).

On the other hand, the minimum correlations of 0.74 and 0.46 for the pathological and normal group respectively reflect that jitter and shimmer are not assessing identical aspects of signal irregularity. While variations of correlation coefficients may in principle be – to some extent – due to chance if a great number of correlations is calculated, the significant difference between intra-jitter/intra-shimmer correlations and the {j3,s15} correlation can also be interpreted in that this particular combination of jitter and shimmer measures gives a significant advantage over the use of just one jitter or shimmer measure or a combination of two jitter or two shimmer measures. This interpretation is in accordance with the variety of findings described in the literature on which grounds no preference can be given to either jitter or shimmer over the other.

[2] {j3,s15} will also be determined as the best combination of jitter and shimmer measures by mutual information analysis in section III C. Therefore only differences between correlation coefficients with regard to this specific pair of measures were tested for their significance in order to keep the number of tests down to a manageable size.

Table III: Rank correlation coefficients of different GNE (log), NNE, and CHNR features for the pathological group and the normal group (for the abbreviations of the features see Table II). Insignificant correlations are marked by an asterisk (*).

	pathological group								normal group							
	gne2	gne3	nne1	nne2	nne3	chnr1	chnr2	chnr3	gne2	gne3	nne1	nne2	nne3	chnr1	chnr2	chnr3
gne1	0.95	0.89	0.53	0.53	0.79	-0.68	-0.71	-0.86	0.81	0.67	-0.09*	-0.12*	0.33*	-0.09*	-0.08*	-0.48
gne2		0.96	0.49	0.50	0.78	-0.64	-0.67	-0.85		0.89	-0.24*	-0.21*	0.27*	0.06*	0.02*	-0.44
gne3			0.45	0.45	0.75	-0.60	-0.62	-0.82			-0.27*	-0.22*	0.14*	0.14*	0.10*	-0.31*
nne1				0.93	0.76	-0.91	-0.86	-0.69				0.86	0.60	-0.91	-0.79	-0.44
nne2					0.72	-0.83	-0.88	-0.64					0.46	-0.73	-0.86	-0.28*
nne3						-0.83	-0.83	-0.96						-0.71	-0.65	-0.94
chnr1							0.94	0.83							0.85	0.66
chnr2								0.83								0.58

2 Correlations between NNE, CHNR, and GNE

The correlation coefficients between GNE, NNE, and CHNR features are given in Table III. For pathological voices the correlation between the different GNE measures is very high (0.89 to 0.96). The smallest value is found for the combination {gne1,gne3} which represents the largest difference in bandwidth.

Of the NNE and CHNR measures, the ones defined for the high frequency region (nne3 and chnr3) are found to be the ones most similar to GNE (correlation 0.75 to 0.79 for nne3, 0.82 to 0.86 for chnr3). Correlations of the other NNE and CHNR measures with GNE are considerably lower (0.45 to 0.53 for nne1/nne2, 0.60 to 0.71 for chnr1/chnr2). The importance of the higher frequency region in the description of pathological voices has already been pointed out by researchers using the NNE (Kasuya et al., 1986b).

If the same frequency ranges are compared, the correlations between CHNR and NNE measures are very high (0.88 to 0.96). Taking into account the design of the two features that both relate the harmonic energy to the non-harmonic energy it seems plausible to interpret this co-variance as indication that CHNR and NNE measure similar voice properties. The correlations also indicate that the choice of the frequency range has a greater effect than the specific method used (i.e., NNE or CHNR).

For the normal voices all NNE and CHNR measures are significantly correlated with the one exception of {chnr3,nne2}. However, no significant correlations are found between GNE measures and NNE/CHNR (with the exceptions of {chnr3/gne1} and {chnr3/gne2}). While one possible explanation may be the limited value range for the normal voices, these findings can also be interpreted as indication that for normal voices GNE is sensitive to other voice characteristics than NNE and CHNR.

3 Correlations of jitter and shimmer with NNE, CHNR, and GNE

Table IV shows the correlations of GNE, NNE, and CHNR with jitter and shimmer. For the pathological group all values are significant. The correlations between jitter/shimmer and GNE are smaller than between jitter/shimmer and NNE/CHNR. The minimum correlation of any "noise measure" with jitter or shimmer measures is found for gne3. All pairwise differences between the correlation coefficients of any of the 9 "noise measures" with both j3 or s15 were tested for their significance (i.e., all combinations of values in the j3 column and in the s15 column for the pathological voice group in Table IV). It was found that correlations of gne2 or gne3 were significantly smaller than correlations of CHNR or NNE measures.

For the normal group no significant correlation between any GNE measure and any jitter or shimmer measure is found. On the other hand, all CHNR and NNE measures are correlated significantly with jitter and shimmer. The insignificant correlations between GNE and the aperiodicity measures for the normal group and the significantly lower correlation for the pathologic group, together with the significant correlations of NNE/CHNR measures with jitter or shimmer can be interpreted that GNE measures additive noise independently of jitter and shimmer. This is in accordance with results on synthetic signals where gne3 was found to be independent of jitter and shimmer (Michaelis et al., 1997a).

The observation that the correlation of GNE with jitter and shimmer only occurs for pathologic voices can be interpreted as follows: GNE on the one hand and jitter and shimmer on the other hand measure two different voice qualities that often appear together in pathological voices (as has been argued by e.g. Eskenazi et al. (1990); Dejonckere (1995)). Jitter and shimmer have been associated with irregularities of oscillation that may be due to morphological changes of tissue properties (e.g., changes of the oscillating masses for tumors or cysts, changes of the elastic properties for vocal fold paralysis) (Lieberman, 1963; Dejonckere, 1995). Additive noise may be attributed to air passing through a glottal leak (e.g., for tumors, vocal fold paralyses, post-operative conditions for tumor patients) (Kasuya et al., 1986a; Dejonckere, 1995). Under these assumptions it is reasonable to expect measures of aperiodicity and noise content to covary for pathological voices (Lieberman, 1963; Dejonckere, 1995). Since voice quality of normal voices is generally regarded to possess many independent degrees of freedom this covariance is not to be expected for the normal group.

Table IV: Rank correlation coefficients of different GNE (log), NNE, and CHNR features with jitter and shimmer for the pathological and the normal group (for the abbreviations of the features see Table II). Insignificant correlations are marked by an asterisk (*).

	j2	j3	j5	j7	j11	j15	s2	s3	s5	s7	s11	s15
pathological group												
gne1	0.66	0.65	0.66	0.66	0.65	0.64	0.60	0.60	0.60	0.60	0.58	0.57
gne2	0.62	0.62	0.63	0.63	0.61	0.59	0.55	0.56	0.56	0.56	0.54	0.53
gne3	0.58	0.58	0.59	0.59	0.57	0.55	0.52	0.53	0.53	0.52	0.50	0.49
nne1	0.82	0.77	0.81	0.83	0.83	0.82	0.85	0.84	0.85	0.86	0.85	0.83
nne2	0.79	0.74	0.78	0.81	0.83	0.83	0.84	0.81	0.83	0.84	0.85	0.84
nne3	0.81	0.78	0.81	0.82	0.80	0.77	0.72	0.71	0.72	0.72	0.71	0.70
chnr1	-0.84	-0.80	-0.84	-0.84	-0.84	-0.82	-0.87	-0.86	-0.88	-0.87	-0.85	-0.83
chnr2	-0.81	-0.78	-0.82	-0.83	-0.84	-0.83	-0.86	-0.85	-0.86	-0.86	-0.85	-0.83
chnr3	-0.79	-0.77	-0.79	-0.79	-0.77	-0.75	-0.70	-0.70	-0.71	-0.70	-0.69	-0.67
normal group												
gne1	0.11*	0.14*	0.10*	0.07*	0.00*	-0.02*	0.15*	0.21*	0.17*	0.11*	0.05*	0.03*
gne2	0.00*	0.05*	0.02*	-0.03*	-0.11*	-0.14*	0.00*	0.07*	0.02*	-0.03*	-0.09*	-0.12*
gne3	-0.09*	-0.05*	-0.08*	-0.13*	-0.20*	-0.22*	-0.05*	-0.00*	-0.04*	-0.08*	-0.13*	-0.17*
nne1	0.59	0.48	0.60	0.66	0.68	0.64	0.71	0.63	0.69	0.73	0.73	0.71
nne2	0.57	0.44	0.58	0.64	0.68	0.66	0.72	0.62	0.70	0.74	0.76	0.74
nne3	0.70	0.65	0.72	0.72	0.67	0.61	0.45	0.43	0.43	0.44	0.45	0.46
chnr1	-0.62	-0.56	-0.64	-0.68	-0.67	-0.62	-0.73	-0.70	-0.73	-0.74	-0.71	-0.69
chnr2	-0.62	-0.53	-0.66	-0.70	-0.69	-0.65	-0.74	-0.70	-0.74	-0.76	-0.74	-0.72
chnr3	-0.60	-0.58	-0.62	-0.61	-0.53	-0.47	-0.39	-0.41	-0.39	-0.38	-0.35	-0.36

4 Correlations of MWC

The correlations of MWC with the other acoustic measures are stated in Table V. The left two columns show the correlation coefficients with jitter and shimmer. The high values for the pathological group (0.78 to 0.89) can be interpreted as indication that MWC measures indeed one aspect of aperiodicity. For the normal group the only insignificant correlation is with j3 which again suggests to use both MWC and this specific jitter measure.

The correlations of MWC with GNE, NNE, and CHNR are stated in the right column. For the pathological group differences between MWC and the 9 "noise measures" were also tested for their significance. It was found that MWC correlates significantly less with GNE (0.58 to 0.65) than with NNE/CHNR (0.77 to 0.95), while for normal voices there is no significant correlation for MWC with GNE measures. In contrast, MWC is significantly correlated with NNE/CHNR (0.43 to 0.80) for normal voices. This can be interpreted that for normal voices deviations from the periodic structure that are measured by MWC are mainly due to variations in shape and period length. If they resulted primarily from additive noise, correlations should be found between MWC and the noise-sensitive GNE measures.

In summary, the correlations of jitter, shimmer, or MWC with GNE are generally either insignificant of significantly smaller than the ones with NNE or CHNR measures. Comparing the different GNE measures, the lowest correlation with different "aperiodicity measures" is found for gne3. We conclude that rank order correlation points to gne3 as the best supplement to the three features jitter, shimmer, and MWC and, more specifically, to the combination {j3,s15,MWC,gne3}.

Table V: Rank correlation coefficients of MWC with the other acoustic features for the pathological and the normal group (for the abbreviations see Table II). Insignificant correlations are marked by an asterisk (*).

	MWC			MWC			MWC	
	pathol.	normal		pathol.	normal		pathol.	normal
j2	0.81	0.41	s2	0.88	0.65	gne1	0.65	0.06*
j3	0.78	0.34*	s3	0.87	0.59	gne2	0.61	-0.07*
j5	0.81	0.42	s5	0.89	0.65	gne3	0.58	-0.12*
j7	0.82	0.46	s7	0.88	0.67	nne1	0.91	0.78
j11	0.82	0.46	s11	0.86	0.63	nne2	0.82	0.61
j15	0.81	0.43	s15	0.84	0.59	nne3	0.77	0.48
						chnr1	-0.95	-0.80
						chnr2	-0.87	-0.66
						chnr3	-0.77	-0.43

C Mutual information

A non-linear approach to find the best feature combination in the description of pathological voices is taken from information theory. According to this approach, the one feature that adds the least mutual information to a given set is regarded as the optimal supplement. For the analyses in this section the data values were replaced by their ranks.

Under the assumption that any irregularity in pathological voices may be found in the F_0 sequence, the energy sequence, or the waveform shape of the glottal cycles (Klingholz, 1987), the best choice of jitter and shimmer measures from our list was determined. This was achieved by calculating the three-dimensional mutual information for all combinations of one jitter measure (j2, j3, j5, j7, j11, j15), one shimmer measure (s2, s3, s5, s7, s11, s15), and MWC. The lowest mutual information was found for the combination {MWC, j3, s15}.

Next, the "noise measure" of our list best supplementing this three-feature set was to be determined. Therefore in this second step the four-dimensional normalized additional information (ΔI_R) was calculated for all combinations of the set

{MWC, j3, s15} plus one of the GNE, NNE, or CHNR measures. Also, a sequence of random numbers (referred to simply as "noise" in the following) was used as fourth "feature" to supply a reference. [3] Theoretically, the predictability of such a "feature" should be zero (corresponding to the highest normalized additional information possible, i.e. ΔI_R=1). [4] Theory predicts that the addition of one of the measures already present should amount to a zero increase of information. Therefore j3 was used as yet another fourth feature to test the performance of the algorithm.

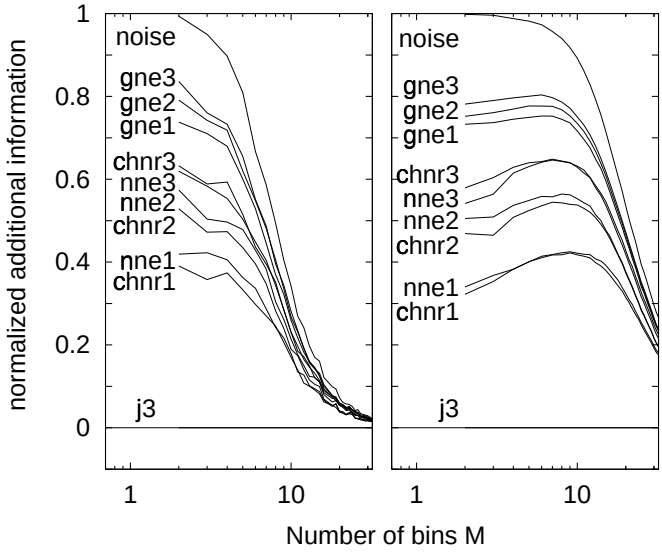

Figure 1: The normalized additional information ΔI_R for different acoustic features as function of the number of bins M per axis (left: pathological group, $n = 447$; right: large pathological group, $n = 13414$).

The results for the pathological group are displayed in Fig. 1 (left). The general drop of the curves reflects the rapid decrease of the average number of data points per hyper-bin when the number of bins per feature axis is increased. For this reason the upper limit of ΔI_R (realized by using noise as fourth feature) has already dropped to approximately 0.95 for M=4. This indicates that in this case ΔI_R can be stated only with a 5% accuracy while 4 bins per axis still give a relatively crude approximation of the probability density of the corresponding measure.

For this reason the same analysis was performed again using the large pathological group (n=13414). Now the drop of the curves starts at higher values of M (Fig. 1, right). In this case the curves reach maximum values for M=8. At this point the value for noise is still greater than 0.95. This signifies that the number of data points per hyper-bin is still sufficiently large to yield dependable results while 8 bins per axis give a much better approximation of the probability distribution than the previous 4 bins. In any case, the ranking of the features with respect to the normalized additional information is not affected by the choice of M and is the same for either data group. As conclusion, the best additional measure indicated by the curves in Fig. 1 is gne3.

Summarizing the results of this section: generally, the ranking of the "noise features" corresponds to the results of the correlation analysis of the previous section – the lower the correlation with the "aperiodicity measures", the higher the normalized additional information. Analysis of the normalized additional information consistently shows for both data sets of pathological voices that about 80% of the possible increase of information can be obtained by gne3 supplementing the set {MWC, j3, s15}. This is about 20% more than the informational gain realized by adding any NNE or CHNR measure. Since GNE gives interpretable information about a voice (in contrast to random noise), the results of this section point to {MWC, j3, s15, gne3} as optimal four-item set out of the given 22 acoustic measures.

D Principal components analysis using singular value decomposition (SVD)

1 Analysis of four-dimensional feature spaces

In the previous section the combination of {MWC, j3,s15} was determined as the best choice of the list of "aperiodicity features" from an information theory perspective. Now principal components analysis is applied to the four-dimensional data space defined by this three-feature set and one additional "noise measure" to determine the underlying dimensionality of the data space.

The results for the three data groups (normal, small and large pathological) are listed in Table VI. The values for the pathological groups show that gne3 consistently maximizes the variance of the second principal axis compared to the NNE and CHNR measures. In particular, the variance of the second principal axis is higher for gne3 (16%) than for chnr3 (9%) and nne3 (10%).

Generally, for the pathological groups only 5 to 8% of the variance are located outside the plane defined by the first two axes. Therefore the informational loss by projecting the four-dimensional space onto a two-dimensional sub-space is relatively small. The statistical congruence of the small and large pathological group is again confirmed by the observation that the differences in the results for both groups never exceed 1%.

For the normal group the four-feature space cannot be projected onto a two-dimensional sub-space without considerable loss of information. A possible interpretation is that for normal voices the four measures vary independently and that a higher-dimensional description is needed for an adequate classification of voice quality.

[3] Two-dimensional plots for different sets of random numbers generated by the internal random number algorithm of the computer exhibited clustered structures. These structures indicated that a uniform distribution using this random number generator was not realized. A much better way of obtaining uniformly distributed values was by taking the sample values of a recording of a whispered voice.

[4] This illustrates again the restrictions that apply to results of the mutual information analysis. The numerical value of ΔI_R does not allow to draw conclusions as to the "knowledge" gained by the addition of the new "feature". However, once the usefulness of a feature is established by other methods or by theory, ΔI_R represents a "gain of information" in the conventional meaning.

Table VI: Variance accounted for by the first four principal axes for the combination of {j3, s15, MWC} with one of the GNE (log), NNE, and CHNR measures for the pathological group ($n = 447$), the extended pathological group ($n = 13414$), and the normal group ($n = 88$).

	pathological group ($n = 447$)				pathological group ($n = 13414$)				normal group ($n = 88$)			
Par	1	2	3	4	1	2	3	4	1	2	3	4
gne1	0.80	0.12	0.05	0.03	0.79	0.12	0.05	0.03	0.51	0.25	0.16	0.09
gne2	0.78	0.14	0.05	0.03	0.77	0.15	0.05	0.03	0.50	0.26	0.15	0.08
gne3	0.76	0.16	0.05	0.03	0.76	0.16	0.05	0.03	0.51	0.25	0.16	0.08
nne1	0.88	0.06	0.03	0.02	0.89	0.06	0.04	0.01	0.68	0.18	0.09	0.05
nne2	0.87	0.06	0.04	0.03	0.88	0.05	0.04	0.03	0.67	0.18	0.10	0.06
nne3	0.82	0.10	0.05	0.03	0.82	0.10	0.05	0.03	0.63	0.18	0.12	0.08
chnr1	0.89	0.06	0.04	0.01	0.89	0.06	0.04	0.01	0.69	0.17	0.09	0.04
chnr2	0.88	0.06	0.04	0.03	0.88	0.06	0.04	0.02	0.68	0.17	0.09	0.06
chnr3	0.83	0.09	0.05	0.03	0.83	0.09	0.05	0.03	0.60	0.19	0.13	0.07

2 Analysis of the complete feature space

PCA supplies a means to analyze the complete 22-dimensional feature space simultaneously and was therefore used to test whether GNE is indeed the main contributor to the second principal axis. With this motivation SVD was performed for the complete set of the 22 acoustic measures stated in Table II. The variance accounted for by the different principal axes is shown in Fig. 2. A two-dimensional space results for the pathological group and a four-dimensional space for the normal group if principal components accounting for less than 5% of the data are considered negligible. For both groups the first axis is the predominant one, accounting for 80% of the variance for the pathological group and 59% for the normal group.

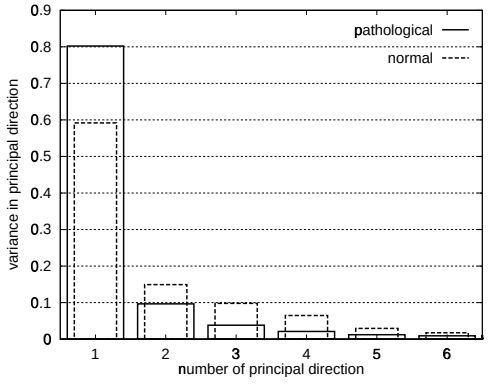

Figure 2: Variance accounted for by the first six principal components extracted from a twenty-two dimensional data space for the normal and pathological group (sorted in descending order).

A deeper insight into the structure of the principal directions is gained by a closer look at the components of the first two principal vectors (Fig. 3). For the pathological group the absolute values of the contribution to the first principal axis are approximately constant (0.2) with the exception of the GNE measures at clearly lower values (Fig. 3, top). For the normal group, where GNE hardly contributes to the first principal axis at all, this difference between GNE and the other measures is even more pronounced. However, the three GNE measures are found to be the main contributors to the second principal axis (Fig. 3, bottom). This indicates GNE as the dominant feature for the data distribution to be two-dimensional instead of just one-dimensional. These findings support the previous results of the SVD on four-feature sub-sets.

E The "hoarseness diagram"

The results of the last section cannot be interpreted easily without a deeper understanding of the analysis methods applied. However, for a clinical application a concise (yet suggestive) way to present the results is desirable. Since the data space of the pathological voices was found to be two-dimensional, a two-dimensional graphical representation seems to be the most obvious way to present the acoustic analysis.

In order to obtain a suitable graph, the coordinate system defined by the first two principal components of the {MWC, j3, s15, gne3}-system is rotated by 63.9 degrees. This leads to a minimization of the gne3-component in the first principal direction (see Table VII, top). Simplifying further, not these exact factors are used but a balanced idealization (Table VII, bottom). An offset of 5 and 1.5 is chosen for the x-axis and the y-axis respectively in order to obtain positive values for all data. The GNE enters the y-coordinate linearly (instead of the logarithmic feature gne3) to supply a better possibility to distinguish between normal and pathologic voices.

The coordinates in this plot of the normalized data are thus given by the equations:

$$\text{x-coordinate} = 5 + \frac{1}{\sqrt{3}}\left(\frac{\text{MWC} + 1.614}{0.574} + \frac{\text{j3} + 0.374}{0.645} + \frac{\text{s15} - 0.757}{0.368}\right) \quad (7)$$

$$\text{y-coordinate} = 1.5 + \frac{0.695 - (1 - 10^{\text{gne3}})}{0.242}. \quad (8)$$

From Eq. (7) it can be seen that only "aperiodicity measures" enter the x-coordinate so that this axis is named *irregularity component*. Eq. (8) shows that the y-axis is based entirely on the "noise feature" GNE. Therefore the ordinate is labeled *noise component*. These two coordinates define the "hoarseness diagram" that allows a quantitative description and graphical interpretation of pathological voice quality (Michaelis et al., 1997b; Fröhlich et al., 1997, 1998a,b).

Table VII: Relative contribution of the four features MWC, j3, s15, and gne3 to the two principal axes rotated by 63.9 degrees. The rotation leads to a minimization of the fourth feature (gne3) in the first principal direction (top: exact values, bottom: balanced idealization used for the "hoarseness diagram").

principal direction	MWC	j3	s15	gne3
1	0.499	0.606	0.619	0.000
2	-0.172	0.078	0.063	0.980
factors used in the "hoarseness diagram":				
1	0.577	0.577	0.577	0.000
2	0	0	0	1

The "hoarseness diagrams" of the large pathological group and the normal group in Fig. 4 reveal that a large area is covered by the pathologic voices. The normal voices are clustered in the lower left region and supply a reference when looking at the undifferentiated pathological group.

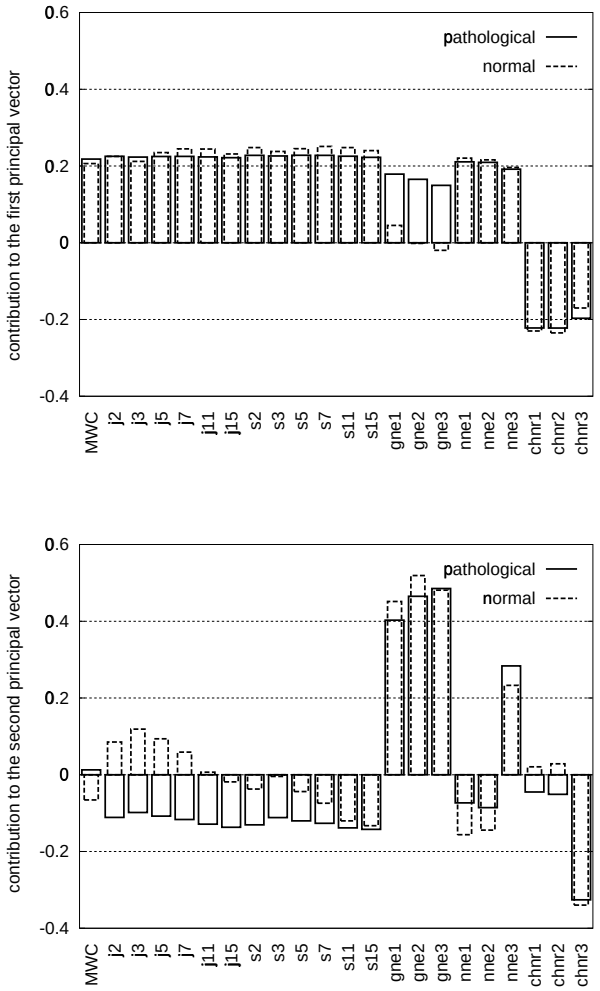

Figure 3: Contribution of the acoustic features to the first and second principal axis obtained by SVD of the 22-dimensional feature space. The variance in the first principal direction (top) is 80% for the pathological group and 59% for the normal group, in the second principal direction (bottom) 9.7% for the pathological group and 15% for the normal group.

IV Conclusion

A two-dimensional description of voice quality was derived from four acoustic voice quality measures. Three features (jitter, shimmer, mean period correlation MWC) describe different aspects of the signal irregularity. Mutual information analysis revealed that the best jitter and shimmer measures chosen from a list of 12 possible ones were the PPQ averaging 3 periods (j3) and the EPQ averaging 15 periods (s15). Results of the PCA suggest that these three measures contribute in roughly equal parts to a common dimension.

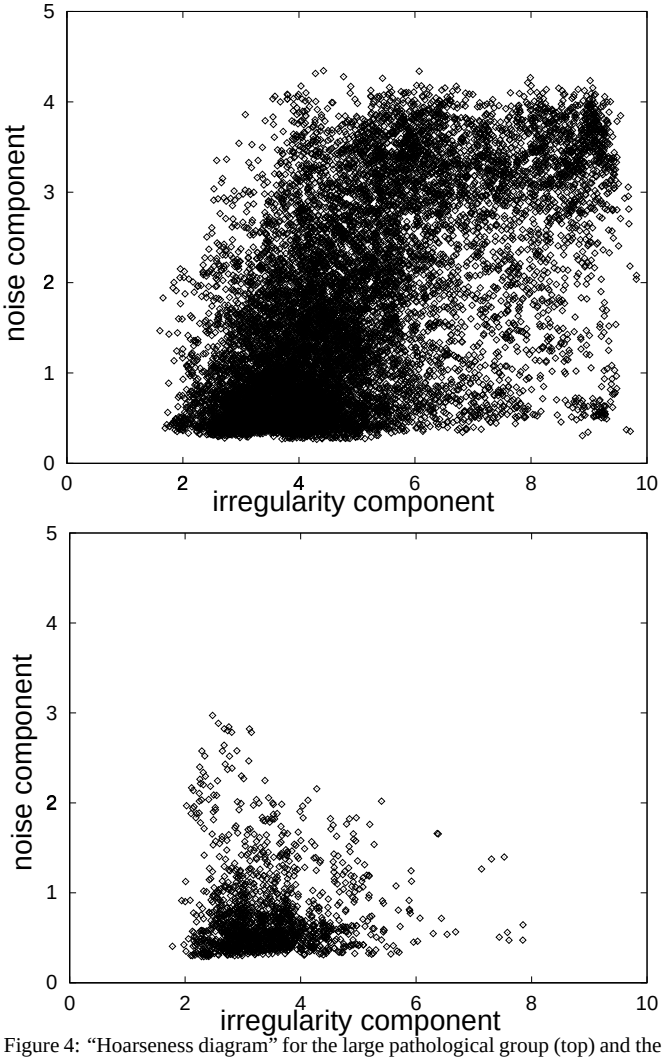

Figure 4: "Hoarseness diagram" for the large pathological group (top) and the normal group (bottom). Each analysis frame is represented by one point with the coordinates given by Eqs. (7), (8).

An additional feature supplying new information about a pathological voice was determined from a list of measures capable to describe the noise content of a speech signal. Correlation analysis, mutual information analysis and principal components analysis consistently showed GNE with a 3kHz bandwidth of the Hilbert envelopes as the best additional measure. Compared to CHNR and NNE, GNE correlated less with jitter and shimmer for pathologic voices and not significantly for normal voices. Therefore GNE should be given preference for the independent measurement of additive noise.

The four features found to give the best description of a pathological voice were used to define the "hoarseness dia-

gram". In this diagram jitter, shimmer, and period correlation contribute in equal parts to the x-coordinate while a linear function of the GNE defines the y-coordinate. The "hoarseness diagram" has already proven to be valuable in differentiating between various phonation mechanisms and specific vocal pathologies (Michaelis et al., 1997b; Fröhlich et al., 1998a,b) as well as in monitoring the progress of voices during voice rehabilitation (Fröhlich et al., 1998a,b). Its correlation to perceptual voice qualities has been shown (Zwirner et al., 1998).

Appendix

The glottal to noise excitation ratio (GNE) is designed as an acoustic measure to assess noise in a pulse train that is typically generated by the oscillation of the vocal folds (Michaelis and Strube, 1995; Michaelis et al., 1997a). It is based on the assumption that glottal pulses resulting from the collision of the vocal folds lead to a synchronous excitation of different frequency bands. Turbulent noise generated at a constriction, on the other hand, leads to an uncorrelated excitation. The synchronism is expressed by the correlations between envelopes calculated for the different frequency bands.

The algorithm to calculate the GNE (more precisely to calculate gne3 according to Table II) looks as follows (for a more detailed description see Michaelis et al. 1997a): 1) linear-predictive inverse filtering of the speech signal to obtain glottal pulses (if present); 2) bandpass filtering of the residual signal in the frequency domain by applying Hanning windows (3000Hz width) at different center frequencies; 3) calculation of the Hilbert envelopes for each frequency band in the frequency domain and back-transformation to the time domain; 4) calculation of correlation coefficients between the different Hilbert envelopes for lags in the range -0.3ms$\leq$t$\leq$0.3ms (difference of center frequencies has to be at least 1500Hz); 5) maximum correlation coefficient defines the GNE.

GNE is sensitive to broad band noise which was tested in a study using synthetic signals (Michaelis et al., 1997a). It reaches its maximum value of 1.0 if the envelopes in two different frequency bands are exactly the same.

However, a general restriction for the generalization of results on synthetic signals lies in the appropriateness of the model used to generate the data. Therefore tests on real speech are equally important. The performance of the GNE has already been discussed for several case studies of various voice conditions (Fröhlich et al., 1997). Generally, for real voices white or high-frequent noise might be present in the speech signal at a level at which it already affects the higher harmonics (due to the spectral tilt) while the lower harmonics still remain relatively unchanged. Since voice signals contain most energy at the low frequencies, in this case the signal periodicity - insofar as it existed - would be hardly affected. However, the correlation between the envelopes in the different bands would be diminished considerably. Therefore, a lower GNE would result.

In such a situation, both an increase of the noise level and a stronger spectral tilt would show up as a further decrease in the GNE. Both aspects have been correlated with breathy voice quality (Hammarberg et al., 1981; Klatt and Klatt, 1990; Hillenbrand et al., 1994; Dejonckere, 1995) which in turn indicates the glottal phonation conditions according to some studies (Fritzell et al., 1983b; Södersten and Lindestad, 1990). Therefore GNE is a promising acoustic feature in the assessment of voice quality with regards to perceptual qualities (Zwirner et al., 1998) or to physiological conditions (Fröhlich et al., 1998a,b).

Acknowledgment

We wish to express our gratitude to Prof. E. Kruse for his encouragement and cooperation. We thank the anonymous reviewers for the many helpful suggestions and comments. Also, we want to thank I. Titze, R. Baken, R. Orlikoff, and J. Sundberg for the discussions of our results. This work is part of a project funded by the Deutsche Forschungsgemeinschaft under Kr 1469/2-1.

References

Arends, N., Povel, D.-J., van Os, E., and Speth, L. (**1990**). "Predicting voice quality of deaf speakers on the basis of glottal charactristics", J. Speech Hear. Res. **33**, 116–122.

Askenfelt, A. G., and Hammarberg, B. (**1986**). "Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures", J. Speech Hear. Res. **29**, 50–64.

Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., and Berke, G. S. (**1996**). "Comparison of Voice Analysis Systems for Perturbation Measurement", J. Speech Hear. Res. **39**, 126–134.

Childers, D., and Lee, C. (**1991**). "Vocal quality factors: Analysis, synthesis, and perception", J. Acoust. Soc. Am. **90**, 2394–2410.

de Krom, G. (**1993**). "A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals", J. Speech Hear. Res. **36**, 224–266.

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

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

Eskenazi, L., Childers, D., and Hicks, D. (**1990**). "Acoustic Correlates of Vocal Quality", J. Speech Hear. Res. **33**, 298–306.

Fraser, A. M., and Swinney, H. L. (**1986**). "Independent coordinates for strange attractors from mutual information", Physical Review A **33**, 1134–1140.

Fritzell, B., Gauffin, J., Hammarberg, B., Karlsson, I., and Sundberg, J. (**1983a**). "Measuring insufficient vocal fold closure during phonation", STL-QPSR **4**, 50–59.

Fritzell, B., Hammarberg, B., Gauffin, J., and Sundberg, J. (**1983b**). "Acoustic inverse filtering of breathy phonation", XIX Congress of the International Association of Logopaedics and Phoniatrics .

Fröhlich, M., Michaelis, D., Strube, H. W., and Kruse, E. (**1997**). "Acoustic voice quality description: Case studies for different regions of the hoarseness diagram". In *Advances in Quantitative Laryngoscopy, 2nd 'Round Table'*, edited by T. Wittenberg, P. Mergell, M. Tigges, and U. Eysholdt (Erlangen, 1997), pp. 143–150.

Fröhlich, M., Michaelis, D., Strube, H. W., and Kruse, E. (**1998a**). "Stimmgütebeschreibung mit Hilfe des Heiserkeits-Diagramms: Untersuchung verschiedener pathologischer Gruppen". In *Aktuelle phoniatrisch-pädaudiologische Aspekte 1997*, edited by M. Gross (Median Verlag, Heidelberg, 1998), Vol. 5. (in press).

Fröhlich, M., Michaelis, D., Strube, H. W., and Kruse, E. (**1998b**). "Voice quality assessment by means of the hoarseness diagram". submitted to J. Speech Lang. Hear. Res., 1998.

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

Hammarberg, B., Fritzell, B., and Schiratzki, H. (**1981**). "Teflon injection in 16 patients with paralytic dysphonia - perceptual and acoustic evaluations", STL-QPSR 1/1981 , 38–57.

Hillenbrand, J., and Houde, R. A. (**1996**). "Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech", J. Speech Hear. Res. **39**, 311–321.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic Correlates of Breathy Vocal Quality", J. Speech Hear. Res. **37**, 769–778.

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

Hillenbrand, J. (**1988**). "Perception of aperiodicities in synthetically generated voices", J. Acoust. Soc. Am. **83**, 2361–2371.

Hirano, M., Hibi, S., Terasawa, R., and Fujiu, M. (**1986**). "Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia", Journal of Phonetics **14**, 445–456.

Hirano, M., Hibi, S., Yoshida, T., Hirade, Y., Kasuya, H., and Kikuchi, Y. (**1988**). "Acoustic Analysis of Pathological Voice", Acta Otolaryngol **105**, 432–438.

Hollien, H., Michel, J., and Doherty, E. (**1975**). "A Method for Analyzing Vocal Jitter in Sustained Phonation", J. Phonetics **1**, 85–91.

Holm, S. (**1979**). "A simple sequentially rejective multiple test procedure", Scand. J. Statist. **6**, 65–70.

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

Kasuya, H., Ogawa, S., Mashima, K., and Ebihara, S. (**1986b**). "Normalized noise energy as an acoustic measure to evaluate pathologic voice", J. Acoust. Soc. Am. **80**, 1329–1334.

Kasuya, H., Endo, Y., and Saliu, S. (**1993**). "Novel acoustic measurements of jitter and shimmer characteristics from pathological voice". In *Eurospeech '93* (1993), Vol. 3, pp. 1973–1976.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers", J. Acoust. Soc. Am. **87**, 820–857.

Klingholz, F. (**1987**). "The measurement of the signal-to-noise ratio (SNR) in continuous speech", Speech Communication **6**, 15–26.

Koike, Y. (**1971**). "Application of Some Acoustic Measures for the Evaluation of Laryngeal Dysfunction", Studia Phonologica (Kyoto University) **7**, 45–50.

Kreiman, J., and Gerratt, B. R. (**1994**). "The multidimensional nature of pathologic vocal quality", J. Acoust. Soc. Am. **96**, 1291–1302.

Kreiman, J., and Gerratt, B. R. (**1996**). "The perceptual structure of pathologic voice quality", J. Acoust. Soc. Am. **100**, 1787–1795.

Kreiman, J., Gerratt, B. R., Precoda, K., and Berke, G. S. (**1992**). "Individual Differences in Voice Quality Perception", J. Speech Hear. Res. **35**, 512–520.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). "Perceptual Evaluation of Voice Quality: Review, Tutorial, and a Framework for Future Research", J. Speech Hear. Res. **36**, 21–40.

Kumar, A., and Mullik, S. (**1996**). "Nonlinear dynamical analysis of speech", J. Acoust. Soc. Am. **100**, 615–629.

Lieberman, P. (**1963**). "Some Acoustic Measures of the Fundamental Periodicity of Normal and Pathologic Larynges", J. Acoust. Soc. Am. **35**, 344–353.

Michaelis, D., and Strube, H. W. (**1995**). "Empirical study to test the independence of different acoustic voice parameters on a large voice database". In *Eurospeech '95*, edited by J. M. Pardo, E. Enriquez, J. Ortega, J.Ferreiros, J. Macias, and F. J. Valverde (1995), Vol. 3, pp. 1891–1894.

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

Michaelis, D., Strube, H. W., and Kruse, E. (**1997b**). "Reliabilität und Validität des Heiserkeits-Diagramms". In *Aktuelle phoniatrisch-pädaudiologische Aspekte 1996*, edited by M. Gross, and U. Eysholdt (Median Verlag, Heidelberg, 1997), Vol. 4, pp. 25–26.

Milenkovic, P. (**1987**). "Least mean square measures of voice perturbation", J. Speech Hear. Res. **30**, 529–538.

Murry, T., Singh, S., and Sargent, M. (**1977**). "Multidimensional classification of abnormal voice qualities", J. Acoust. Soc. Am. **61**, 1630–1635.

Pinto, N. B., and Titze, I. R. (**1990**). "Unification of perturbation measures in speech signals", J. Acoust. Soc. Am. **87**, 1278–1289.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1989**). *Numerical Recipes in C* (Cambridge University Press, 1989).

Qi, Y., and Hillman, R. E. (**1997**). "Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals", J. Acoust. Soc. Am. **102**, 537–543.

Rammage, L. A., Peppard, R. C., and Bless, D. M. (**1992**). "Aerodynamic, Laryngoscopic, and Perceptual-Acoustic Characteristics in Dysphonic Females with Posterior Glottal Chinks: A Retrospective Study", Journal of Voice **6**, 64–78.

Schoentgen, J., and de Guchteneere, R. (**1997**). "Predictable and random components of jitter", Speech Comm. **21**, 255–272.

Södersten, M., and Lindestad, P.-Å. (**1990**). "Glottal closure and perceived breathiness during phonation in normally speaking subjects", J. Speech Hear. Res. **33**, 601–611.

Titze, I. R., and Liang, H. (**1992**). "Comparison of f0 extraction methods for high precision voice perturbation measurement", NCVS Status and Progress Report **3**, 97–115.

van den Brink, W., and Koele, P. (**1987**). *Statistiek, deel 3, Toepassingen* (Boom Meppel Amsterdam, 1987).

Yumoto, E., Gould, W. J., and Baer, T. (**1982**). "Harmonic-to-noise ratio as an index of the degree of hoarseness", J. Acoust. Soc. Am. **71**, 1544–1550.

Zwirner, P., Michaelis, D., Fröhlich, M., Strube, H. W., and Kruse, E. (**1998**). "Korrelationen zwischen perzeptueller Beurteilung von Stimmen nach dem RBH-System und akustischen Parametern". In *Aktuelle phoniatrisch-pädaudiologische Aspekte 1997*, edited by M. Gross (Renate Gross Verlag, Berlin, 1998), Vol. 5. (in press).