ACUSTICA · acta acustica
Vol. 83 (1997) 700 – 706

Glottal-to-Noise Excitation Ratio – a New Measure for Describing Pathological Voices

D. Michaelis, T. Gramss, H.W. Strube
Drittes Physikalisches Institut, Georg-August-Universität, Bürgerstr. 42-44, D-37073 Göttingen, Germany, strube@physik3.gwdg.de

Dedicated to Manfred R. Schroeder on occasion of his 70th birthday

Summary
In this article a new acoustic parameter for the objective description of voice quality is introduced. It is based on the correlation coefficient for Hilbert envelopes of different frequency bands. The parameter indicates whether a given voice signal originates from vibrations of the vocal folds or from turbulent noise generated in the vocal tract and is thus related to (but not a direct measure of) breathiness. Therefore it is named Glottal-to-Noise Excitation Ratio (GNE Ratio). GNE is compared to HNR (Harmonics-to-Noise Ratio) and NNE (Normalized Noise Energy), existing measures also sensitive to additive noise (turbulence). Experiments with artificial signals show that only the GNE is almost independent of frequency modulation noise (jitter) and amplitude modulation noise (shimmer).

1. Introduction

In 1961, Lieberman proposed the first acoustic voice parameter in pathological-voice analysis (Lieberman, 1961). Since then, many different parameters have been introduced. They fall into two groups: Parameters describing additive noise (turbulence) such as HNR (Yumoto *et al.*, 1982; de Krom, 1993), NNE (Kasuya *et al.*, 1986), SNR (Klingholz, 1987), N/S (Muta and Baer, 1988), and parameters describing frequency modulation noise (jitter) and amplitude modulation noise (shimmer) like PF and PPQ (Kasuya *et al.*, 1993). These two groups correspond to different ways of exciting the vocal tract: by (possibly irregular) glottal oscillations and by turbulent air-flow noise. Since we want to measure different voice characteristics in a *production*-oriented way, the parameters describing additive noise should be independent of modulation noise. However, this is not the case for parameters introduced so far (de Krom, 1993; Muta and Baer, 1988).

The GNE represents a new approach to quantify the amount of voice excitation by vocal-fold oscillations versus excitation by turbulent noise. It is thus closely related to breathiness, although the latter is a multidimensional perceptual phenomenon which we do not intend to model directly. (Investigations of the relation between GNE and subjective assessment are being carried out.) Methods introduced in the past define parameters for breathiness or additive noise in either the frequency domain or the time domain (de Krom, 1995). They depend on the regularity of several glottal oscillations. The GNE is applicable even for highly irregular glottal oscillations.

In the experiments described in this article, the new parameter GNE is compared with CHNR (Cepstrum based Harmonic to Noise Ratio (de Krom, 1993)) and with NNE. The sensitivity of the parameters to additive noise and to modulation noise (jitter and shimmer) is measured.

As a review, we will first give brief definitions for NNE and HNR as they have been used in this work.

Received 24 May 1996,
accepted 28 October 1996.

2. Definition of NNE and HNR

The definitions for NNE and HNR differ slightly for different authors. We used the following implementation.

NNE is the ratio between the energy of noise and total energy of the signal (both measured in dB). Between the harmonics, the noise energy is directly obtained from the spectrum. Within a harmonic, the noise energy is assumed to be the mean value of both adjacent minima in the spectrum. A harmonic is assumed to have the width of a Fourier transformed Hann time window, in our case $2M/N$ samples, where $M = 7/f_0$ is the Hann window size, f_0 the fundamental frequency, and and N is the number of samples for the Fast Fourier Transform, the smallest power of 2 which is greater than M (Kasuya *et al.*, 1986).

If the harmonics are broadened because of jitter or shimmer, the energy outside the window defined for the harmonic is erroneously assigned to noise energy. As a consequence, the noise measured by NNE appears to be increasing. The authors varied the frequency range to obtain best discrimination between normal and pathological (glottal cancer) voice. They found a frequency range from 1 to 5 kHz for NNE optimal for discrimination.

Roughly speaking, CHNR (this is the cepstrum based HNR (de Krom, 1993)) is the inverse of NNE: It is the ratio between total energy and energy of noise (both measured in dB). However, the energies are obtained in a different way: The procedure to measure the noise is that the cepstral peaks at the fundamental period and its multiples are removed.

Essentially, the spectral energy between the harmonics below the lines that connect the minima is considered to be noise energy. Therefore, the inverse CHNR is generally

ACUSTICA · acta acustica
Vol. 83 (1997)

larger than the NNE. Due to jitter and shimmer, the harmonics are broadened and the minima of the spectrum are less deep. As a consequence, in presence of jitter and shimmer, the noise energy is overestimated by CHNR.

Each experiment with NNE and CHNR has been performed for three frequency ranges: 60 to 2000 Hz, 60 to 5000 Hz, and 1000 to 5000 Hz. Since the results do not depend on the frequency ranges chosen, only the frequency range from 1000 to 5000 Hz for the NNE and from 60 to 2000 Hz for the CHNR is used in the following.

We now proceed by defining GNE.

3. Glottal-to-Noise Excitation Ratio

Our method is based on the correlation between Hilbert envelopes of different frequency channels. Triggered by a single glottis closure, all the frequency channels are simultaneously excited (Figure 1c), so that the envelopes in all channels share the same shape, leading to high correlation between the envelopes. The shape of each excitation pulse is practically independent of preceding or following pulses.

In case of turbulent signals (noise, whisper) a narrowband noise is excited in each frequency channel. These narrow band noises are uncorrelated (if the windows that define adjacent frequency channels do not overlap too much).

In this way it can be achieved that deviations from periodicity do not influence the degree of turbulence measured by GNE.

A loosely related interband correlation technique has been suggested in (Aures, 1985) as a model of roughness perception. Our approach, however, is only motivated by the speech production process and signal theory; it does not intend to model any perceptive effects.

The GNE factor is calculated in the following way (explanations see below):

1) Down-sampling to 10 kHz.
2) Inverse filtering of the speech signal (see Figure 1b).
3) Calculating the Hilbert envelopes (the absolute value of the complex analytic signal as depicted in Figure 1d) of different frequency bands with fixed bandwidth and different center frequencies (see Figure 1c).
4) Consider every pair of envelopes for which the difference of their center frequencies is equal or greater than half the bandwidth: calculate the cross correlation function between such envelopes.
5) Pick the maximum of each correlation function.
6) Pick the maximum from the maxima in 5).

The inverse filtering in step 2 is applied to transform the speech wave (Figure 1a) to a sequence of narrow pulses, approximately delta functions. This is achieved by flattening the spectrum so that the harmonics have about the same amplitude (Figure 1b). The peaks of the pulses presumably indicate the instants of glottal closure. The recovery of the sequence of delta functions from the speech wave is not perfect for voice samples that are digitized with 48

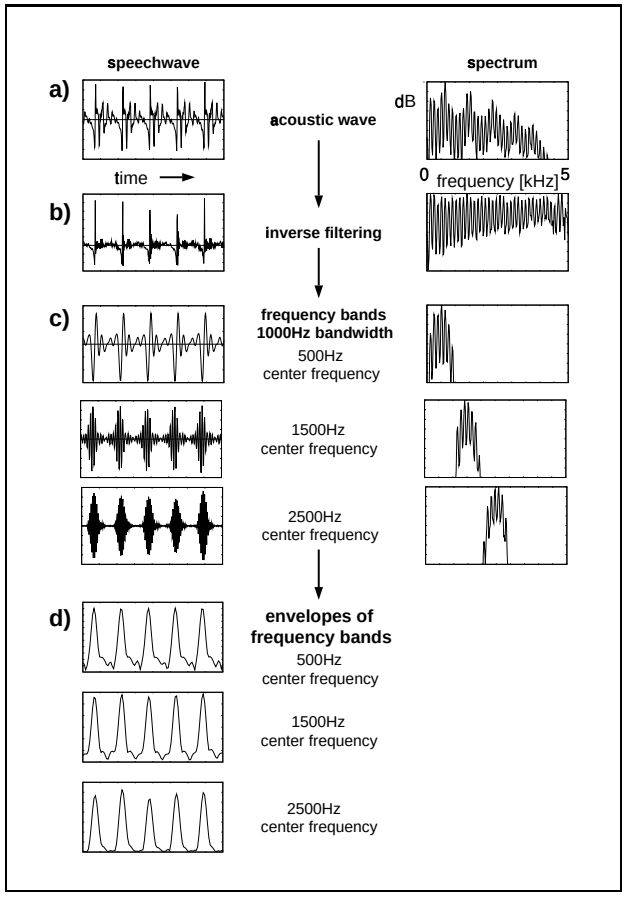

Figure 1. Signal processing for GNE calculation. Left: time functions, right: corresponding logarithmic power spectra

kHz or 50 kHz sampling frequency, because the voice energy nearly vanishes above 5 kHz. Therefore, the signal is first down-sampled to 10 kHz sampling frequency (step 1). The inverse filtering is then done by calculating the linear-prediction error signal, using a predictor of 13th order computed by the "autocorrelation method" with a Hann window of 30 ms length and 10 ms shift between successive frames (Markel and Gray, 1976).

The calculation of the Hilbert envelopes (step 3) is done most efficiently in the frequency domain, without using a filter bank: a) Apply a real discrete Fourier transformation (DFT) on the time signal (Press *et al.*, 1989). (The signal is real, so that the Fourier components at negative frequencies are simply the complex conjugates of those at positive ones. These components at negative frequencies do not have to be calculated in a real DFT.) b) Select a frequency band from the complex spectrum and apply a Hann window. c) Double the length of the signal obtained from b) by padding zeros (that is: setting the values at negative frequencies to zero). d) Apply an inverse Fourier transform. e) Take the absolute value of the complex signal. The steps b) through e) are applied to each frequency band.

In general, the envelopes have different phases. Therefore, it is not sufficient to calculate the zero-time-shift correlation. The reason for the phase shifts might be that the

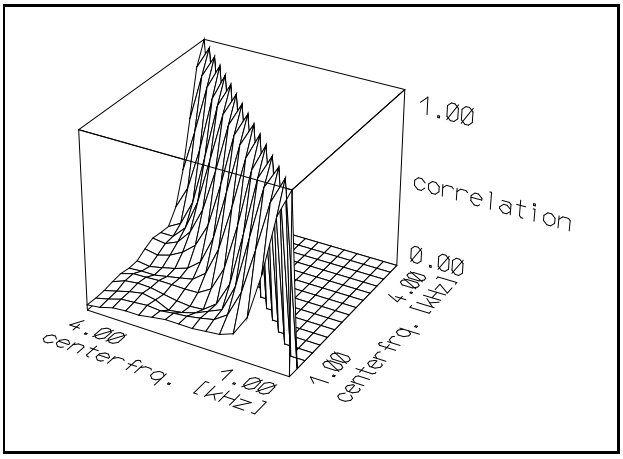

Figure 2. Correlation matrix for a random noise signal (bandwidth of envelopes is 2000 Hz). Correlations between overlapping frequency bands are small if the distance of the center frequencies exceeds half the bandwidth (1000 Hz).

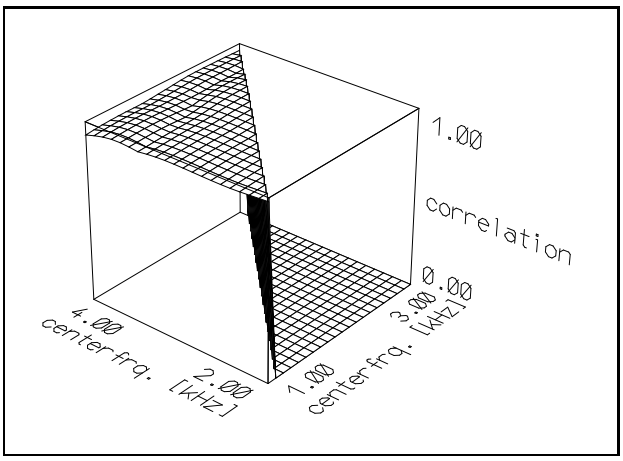

Figure 3. Correlation matrix for a normal voice.

maximum excitation of different frequencies does not occur at exactly the same time during glottal closure (the glottal pulses are not actual delta functions). The delay used for the correlation function between two envelopes in step 4 ranges between -3 and $+3$ samples ($= \pm 0.3$ ms). The maximum within this range is picked for each correlation function (step 5). These maxima are shown in Figures 2 through 4. Finally in step 6, the maximal correlation is chosen as the GNE parameter.

In Figure 2, the results of the procedure applied to white noise is depicted. The envelope correlation is small if the center frequencies differ by at least half the bandwidth.

In Figure 3, the maximal correlation coefficients for a normal (i.e. not pathologically altered) recorded voice sample are shown. Because of the symmetry of the matrix, it is only necessary to compute half the correlation matrix. Therefore, half of the elements are set to zero.

Even for some normal voice a few of the correlations between envelopes are small (Figure 4). In these cases, the harmonic structure vanishes in spectral dips between the

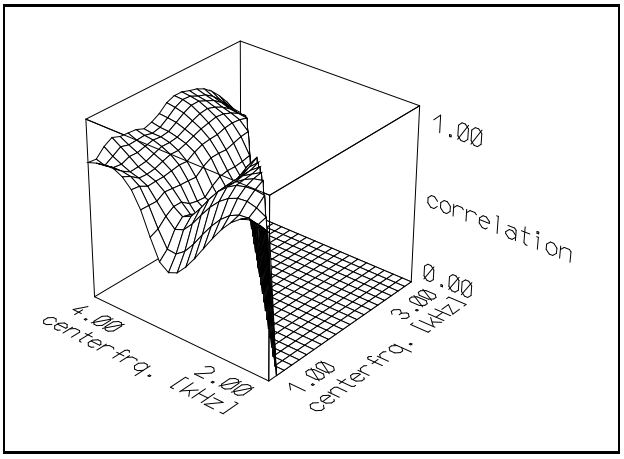

Figure 4. Correlation matrix for a normal voice with several small correlation coefficients.

formants. Thus, this is no indication for a pathologic voice. On the other hand, high correlations are never observed for aphonic voices. Therefore the GNE allows a consistent description of the amount of turbulent excitation in the whole range from normal voices over slightly breathy up to aphonic ones.

4. Analysis of synthesized signals

In this section the influence of additive noise, jitter and shimmer on GNE, CHNR and NNE is examined.

4.1. Test signals

The test signal $s(t)$ is the sum of a sequence of delta functions $D(t)$ and noise $n(t)$: $s(t) = D(t) + n(t)$.

$D(t)$ is a sum of delta functions $\delta(t)$ ($\delta(0) = 1$ and $\delta(t) = 0$ for $t \neq 0$):

$$D(t) = \sum_i (1 + \frac{S}{100\%} r_{1i})\, \delta(t - t_i), \tag{1}$$

$$t_i - t_{i-1} = (1 + \frac{J}{100\%} r_{2i}) T, \tag{2}$$

where r_{1i} and r_{2i} are random numbers from a normal distribution with standard deviation $\sigma = 1$, restricted to ($-3\sigma \leq r_{1i}, r_{2i} \leq 3\sigma$), J and S are the amounts of period and amplitude perturbation (jitter and shimmer) in percent and T is the period length. The component $D(t)$ is generated with 200 kHz sampling frequency to achieve a sufficient time resolution for small jitter values. Then $D(t)$ is lowpass filtered (30th order, 4.5 kHz cutoff frequency) and down-sampled to 10 kHz sampling frequency. The noise component $n(t)$ is a uniformly distributed random number sequence with 10 kHz sampling frequency. The duration of a sequence $s(t)$ is 1s (10000 samples).

ACUSTICA · acta acustica
Vol. 83 (1997)

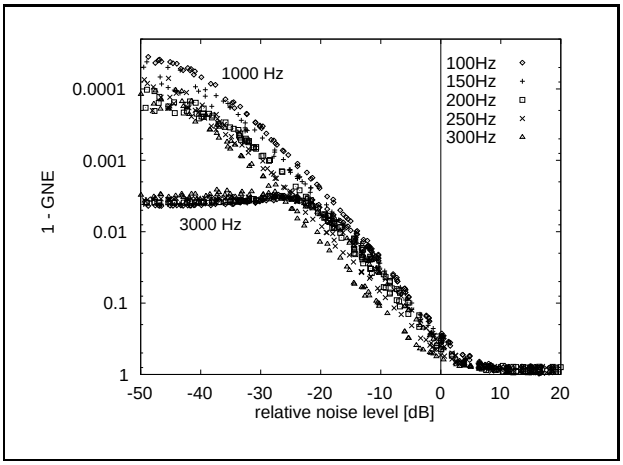

Figure 5. Dependence of GNE on the RMS of the noise component. The envelope bandwidths are 1000 Hz (upper point set) and 3000 Hz (lower set). Here and in the following figures the point symbols denote fundamental frequencies (100 to 300 Hz).

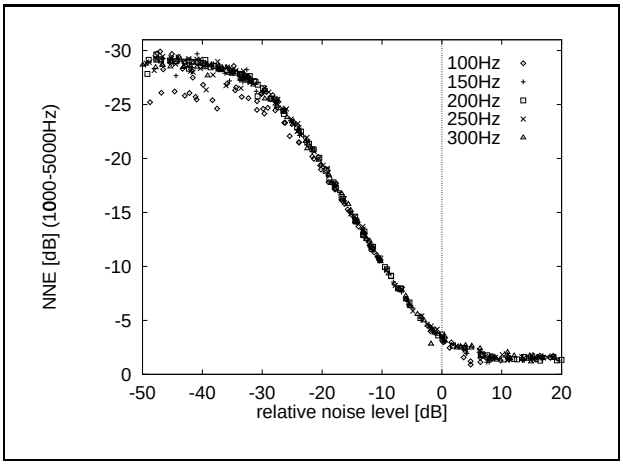

Figure 6. Dependence of NNE on the RMS of the noise component. The frequency ranges from 1000 Hz to 5000 Hz.

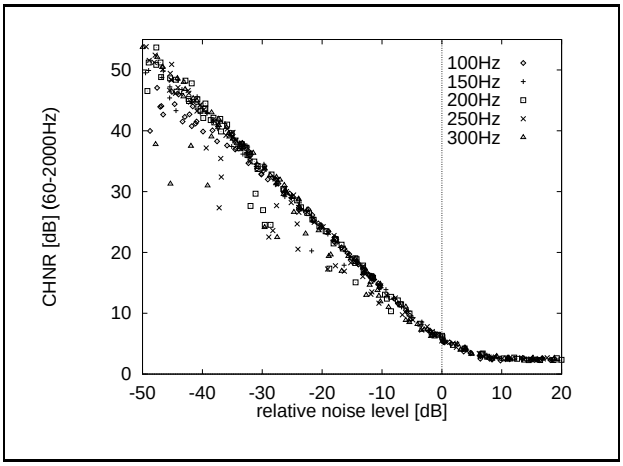

Figure 7. Dependence of CHNR on the RMS of the noise component. The frequency ranges from 60 Hz to 2000 Hz.

4.2. Variations of the noise component

To examine the influence of additive noise on GNE, NNE and CHNR, the Root Mean Square (RMS) of $D(t)$ is set to 1 (0 dB). The signal to noise ratio (RMS of the noise component $n(t)$) is chosen randomly between -50 dB and $+20$ dB (equally distributed on the dB scale). $D(t)$ is generated with different mean fundamental frequencies: 100 Hz, 150 Hz, 200 Hz, 250 Hz, and 300 Hz. The fundamental frequency is randomly varied by $\pm 5\%$. Jitter and shimmer are randomly varied in the range of 0.001% to 0.01% (log. equally distributed). These random variations are introduced to avoid numerical side effects. 1000 sequences were synthesized for each frequency range. For the sake of lucidity, only 100 representative data points are drawn in the following figures, and the ordinate is $1 -$ GNE shown logarithmically downwards (thus greater GNE is higher in the plot).

The GNE was calculated for three different bandwidths of the envelopes (1000 Hz, 2000 Hz, and 3000 Hz). In Figure 5, the GNE in dependence on the noise level is depicted for the bandwidths 1000 Hz (upper point set) and 3000 Hz (lower point set).

For 1000 Hz bandwidth, 51 envelopes were calculated with center frequencies from 500 Hz to 4500 Hz in steps of 80 Hz. The GNE decreases monotonically from 0.9999 to zero as the relative noise level increases from -50 dB to $+20$ dB. An approximately linear decay of $\log(1 - \text{GNE})$ is observed between -5 dB and -35 dB. GNE decreases monotonically with increasing fundamental frequency.

For 2000 Hz bandwidth (not shown, similar to 3000 Hz), 31 envelopes were calculated with center frequencies from 1000 to 4000 Hz in steps of 100 Hz. The GNE decreases from 0.998 to zero. The $\log(1 - \text{GNE})$ is approximately linearly related to the relative noise level in the range from -20 dB to $+5$ dB. GNE also decreases monotonically with increasing fundamental frequency.

For 3000 Hz bandwidth, 21 envelopes were calculated with center frequencies from 1500 to 3500 Hz in 100 Hz steps. The $\log(1 - \text{GNE})$ is again approximately linearly related to the relative noise level in the range from -20 dB to $+5$ dB. The dependence on the fundamental frequency is lower than in the 2000 Hz bandwidth case. The minimal $1 -$ GNE value 0.004 is higher than in the 1000 Hz and 2000 Hz bandwidth cases.

The NNE (Figure 6) and the CHNR (Figure 7) also depend monotonically on the relative noise level (RNL). The NNE increases from -50 dB at -30 dB RNL to about -2 dB at 5dB RNL. The CHNR decreases in the same interval from about 55 dB to about 2 dB. The linear regime can be found from -20 dB to -5 dB RNL for the NNE and from -50 dB to about 5 dB RNL for the CHNR. In contrast to the findings for the GNE, there are relatively many points that strongly deviate from the mainly linear curve.

4.3. Variation of jitter

For one of the experiments, the jitter is randomly chosen between 0.01% and 30% (log. equally distributed) to mea-

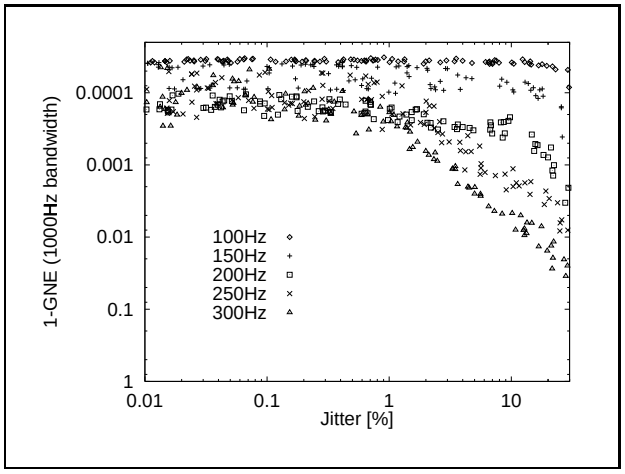

Figure 8. Dependence of GNE on jitter. The envelope bandwidth is 1000 Hz.

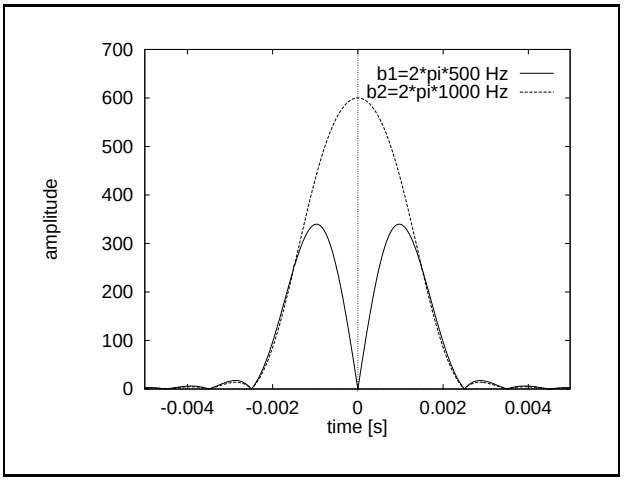

Figure 9. The Hilbert envelope of two delta functions with 1 ms distance in two channels with 500 Hz and 1000 Hz center frequency and 1000 Hz bandwidth.

sure the jitter dependence of GNE, NNE and CHNR. A faint noise was added with a noise level randomly chosen between -50 dB to -49 dB. Again, for the different experiments, the fundamental frequency is randomly varied by $\pm 5\%$ (equally distributed) and shimmer is randomly varied between 0.001% and 0.01% (log. equally distributed).

The jitter dependence of GNE is illustrated in Figure 8. The bandwidth is 1000 Hz. GNE decreases above 1% jitter. The dependence increases with the fundamental frequency. This effect can be easily understood as an interference effect of the Hilbert envelopes from successive delta functions:

The Hilbert envelope $|\sigma(t)|$ of two delta functions separated by a time delay Δt,

$$f(t) = \delta(t + \frac{\Delta t}{2}) + \delta(t - \frac{\Delta t}{2}) \quad , \tag{3}$$

filtered with the bandwidth $a/2\pi$ [Hz] and the center frequency $b/2\pi$ [Hz] reads:

$$|\sigma(t)| = \sqrt{H_{a+}^2 + H_{a-}^2 - 2H_{a+}H_{a-}\sin^2(b\Delta t/2)} \tag{4}$$

with $H_a(t) \equiv (\pi \sin at)/2t(\pi^2 - a^2t^2)$ being the Hilbert envelope of a single bandpass filtered delta function, which is independent of b, $H_{a+} \equiv H_a(t + \frac{\Delta t}{2})$ and $H_{a-} \equiv H_a(t - \frac{\Delta t}{2})$. $|\sigma(t)|$ is dependent on b. $H_a(t)$ is small if at is large, therefore the product $H_{a+}H_{a-}$ is negligible if Δt is sufficiently large to separate H_{a+} and H_{a-}. If Δt is small, the product $H_{a+}H_{a-}$ becomes important and the interference term $2H_{a+}H_{a-}\sin^2 b\frac{\Delta t}{2}$ leads to different envelopes in different frequency bands (see Figure 9). As a consequence, correlations between the envelopes are decreased. The higher the jitter or the fundamental frequency, the more probable are shorter delays between successive delta functions, which result in smaller correlations.

The width of the Hilbert envelope of a single (filtered) delta function $H_a(t)$ depends on the product at only. If the bandwidth is broadened, the envelope and the interference

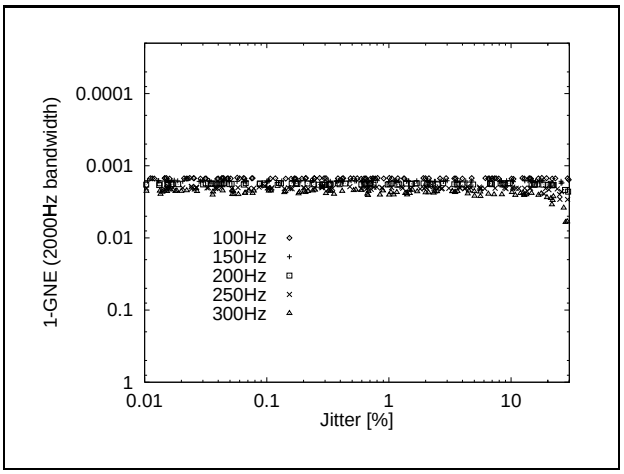

Figure 10. Dependence of GNE on jitter. The envelope bandwidth is 2000 Hz.

effects become smaller. Thus GNE is less sensitive to jitter for higher bandwidths. From Figure 10 it can be seen that the GNE is indeed independent of jitter (and the fundamental frequency) at 2000 Hz bandwidth (and higher).

This is very different for the NNE and the CHNR. Especially the NNE is very much dependent on jitter: At a fundamental frequency of 100 Hz we obtain the same NNE for an experiment with either no noise and one percent jitter (Figure 11) or 5 dB RNL and no jitter (Figure 6). That is, by measuring the NNE, it is not possible to quantify the RNL independently of period perturbations. For the CHNR it is similar: A one percent jitter with no noise present (Figure 12) results in the same CHNR as a RNL of -10 dB and no jitter (Figure 7).

4.4. Variation of shimmer

In order to measure the shimmer dependency of GNE, NNE and CHNR shimmer is randomly chosen between

ACUSTICA · acta acustica
Vol. 83 (1997)

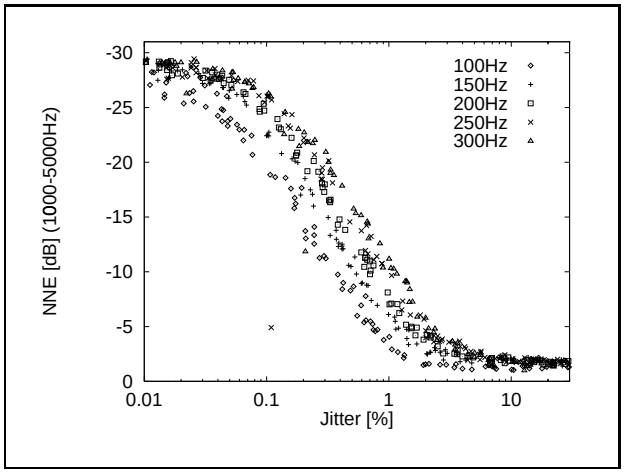

Figure 11. Dependence of NNE on jitter. The frequency ranges from 1000 Hz to 5000 Hz.

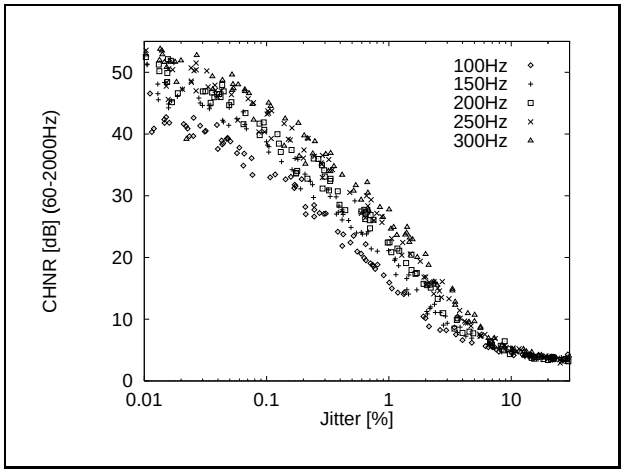

Figure 12. Dependence of CHNR on jitter. The frequency ranges from 60 Hz to 2000 Hz.

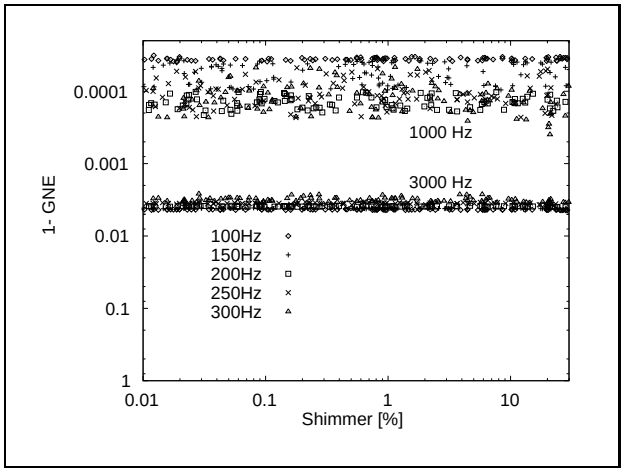

Figure 13. Dependence of GNE on shimmer. The envelope bandwidths are 1000 Hz (upper point set) and 3000 Hz (lower set).

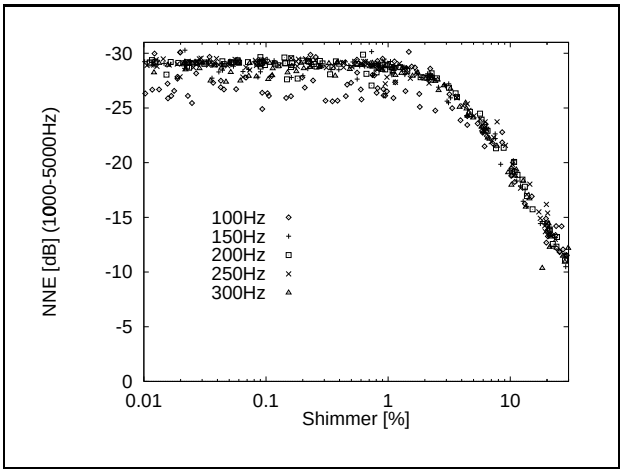

Figure 14. Dependence of NNE on shimmer. The frequency ranges between 1000 Hz and 5000 Hz.

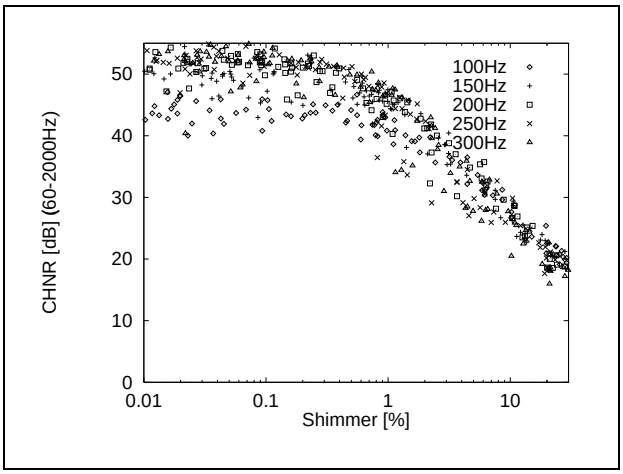

Figure 15. Dependence of CHNR on shimmer. The frequency ranges between 60 Hz and 2000 Hz.

0.01% to 30% (log. equally distributed). Again, faint noise was added with a noise level randomly chosen between -50 dB and -49 dB and the fundamental frequency is varied randomly by $\pm 5\%$ (equally distributed). Jitter is randomly varied between 0.001% and 0.01% (log. equally distributed). The results are shown in Figures 13 to 15. In all the cases, the GNE is independent of shimmer.

Though the dependence on shimmer of NNE and CHNR is less obvious than their dependence on jitter, it can still clearly be observed (see Figures 14 and 15). Above a shimmer of 3%, the NNE increases significantly and the CHNR decreases at shimmer of less than 0.3%.

4.5. CHNR and NNE do not work on vocal fry

As a real-voice example for the dependence of NNE and CHNR on jitter and the independence of GNE of jitter, we compared the three parameters for a normal voice with those for a recording of vocal fry. The spectrum and the Hilbert envelope are shown in Figure 16, together with

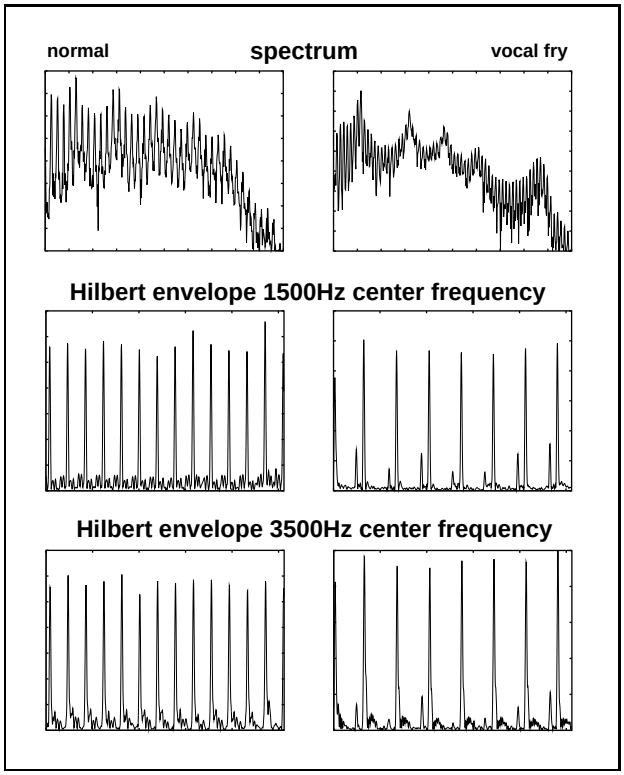

Figure 16. Spectra (abscissa: frequency) and Hilbert envelopes (abscissa: time) for a normal (natural) voice and natural vocal fry each 102.4 ms duration. Normal: GNE (3000 Hz bandwidth) 0.977; NNE (1-5 kHz) -16.65 dB, (60-2000 Hz) -21.03 dB; CHNR (1-5 kHz) 22.95 dB, (60-2000 Hz) 28.24 dB. Vocal fry: GNE 0.970; NNE -1.05 dB, -4.04 dB; CHNR 6.62 dB, 9.83 dB.

the parameters in the figure caption. Significant secondary peaks preceding the main excitation are visible (the instances of glottal opening?) but are also correlated between the bands.

In the spectrum of the vocal fry recording the harmonics are less obvious than in case of normal voice. This is because they are broadened due to jitter. But the Hilbert envelopes are very similar. Therefore the GNE is about the same for both examples, while NNE and CHNR overestimate the noise energy.

5. Conclusion

We have presented a new approach to quantify the amount of voice excitation by glottal oscillations versus excitation by turbulent noise, which ist robust against irregularities of the glottal oscillations. For comparison, the known methods NNE and CHNR have been considered.

In normal voices a jitter up to 1% and shimmer up to 3% can be observed (depending on the method of computation). It has been shown that even in these cases the NNE and CHNR cannot distinguish between variations of amplitude or periodicity and additive noise. As a consequence, a proper evaluation of voice quality cannot be obtained by NNE or CHNR methods. In particular, there is no possibility to reliably distinguish between pathological changes leading to turbulent noise and others leading to irregular glottal excitation.

We introduced a new parameter, the GNE, which is a reliably measure for the relative noise level even in the presence of strong amplitude and periodicity variations. Thus, by regarding jitter, shimmer, and GNE, it is now possible to obtain a more reliable picture of voice quality.

Meanwhile, experiments with numerous subjects with pathological voices have been performed. Results and a discussion of clinical applicability will be published elsewhere. However, it may be said here that these experiments strictly confirm GNE's independence of modulation noise, and a two-dimensional plot with one GNE axis and one combined modulation-noise axis yields a clear classification of various pathologies. The relation of GNE to subjective assessment (for instance, RBH) is being investigated.

Acknowledgement

The authors would like to thank their coworkers at the Abteilung für Phoniatrie und Pädaudiologie, P. Zwirner and D. Hunt as well as the speech therapists, and in particular Prof. Dr. Kruse.

We are especially grateful to the former head of the Drittes Physikalisches Institut, Prof. Dr. M. R. Schroeder, from whom the authors learned more on acoustics and related fields than from any other person.

This work is part of a project funded by the Deutsche Forschungsgemeinschaft under Kr 1469/2-1.

References

Aures, W. (1985). Ein Berechnungsverfahren der Rauhigkeit. *Acustica* **58**, 268–281.

de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J. Speech and Hearing 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 and Hearing Res.* **38**, 794–811.

Kasuya, H., Endo, Y., and Saliu, S. (1993). Novel acoustic measurements of jitter and shimmer characteristics from pathological voice. *EUROSPEECH '93*, Berlin, 1973–1976.

Kasuya, H., Ogawa, S., and Kikuchi, Y. (1986). An adaptive comb filtering method as applied to acoustic analyses of pathological voice. *ICASSP 86*, Tokyo, 669–672.

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

Lieberman, P. (1961). Perturbation in vocal pitch. *J. Acoust. Soc. Am.* **33**, 597–603.

Markel, J. and Gray, A. (1976). *Linear Prediction of Speech*. Springer, Berlin/Heidelberg/New York.

Muta, H. and Baer, T. (1988). A pitch-synchronous analysis of hoarseness in running speech. *J. Acoust. Soc. Am.* **84**, 1292–1301.

Press, W. H. et al. (1989). *Numerical Recipes in C*. Cambridge University Press, Cambridge.

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.