ARCHIVES OF ACOUSTICS
27, 4, 319–329 (2002)

BREAKING A LANCE FOR THE PSYCHOACOUSTICS IN PHONIATRICS

F. KLINGHOLZ

(Former: University of Munich, Faculty of Medicine)
Am Hoehacker 3
D-82229 Seefeld
Germany

Drawbacks of the acoustical as well as the perceptual voice analysis are discussed. The psychoacoustics is proposed as a new framework for a way out of the dilemma of the traditional voice analysis methods. The relevant features of the psychoacoustic measures with reference to the voice are introduced.

1. Introduction

The acoustical voice analysis (AVA) has a relatively long history. However, its usefulness in collecting results and hints for a diagnosis of voice disorders (organic or functional, hyper- or hypofunctional, mass lesion or paralysis, etc.) is questionable. The same appraisal can apply to the outcomes of the perceptive voice analysis (PVA). This paper tries to reveal the sources of this insufficiency by a consideration of the drawbacks of the AVA and the PVA. Moreover, a framework for future methodology is given based on psychoacoustics.

In the second half of the 19^{th} century, the first studies which used AVA were undertaken. Simple devices for analysis were at hand and the laryngeal mirror revealed the physiological activity of the vocal folds. An improvement of technology and the first mathematical models which could be applied to the voice apparatus [17] resulted in progress within the first decades of the 20^{th} century. The analysis concentrated on the exploration of the normal vocal function e.g. to describe the acoustical features of the voice during singing.

In the middle of the 20^{th} century, when electronic devices had reached a higher level of sophistication and were more common, connections between physiology and technical disciplines like communication engineering and technical acoustics were made, and a theoretical framework was available, the AVA was increasingly applied to the pathological voice [13]. A real boom of the AVA started, when microprocessors were available. The run on the AVA was so exaggerated that the physiological base of the measurements and the analysis by auditive perception were truust into the background. This trend can be demonstrated best by a publication in the *Medical Equipment Journal of Japan* [10]. In the article, a system by Ebihara was presented as being able not only to recognize laryngeal cancer acoustically but also to differentiate four stages of the cancer.

On the other hand, there are publications by BAKEN and ORLIKOFF [2] and KLINGHOLZ [8], which question the general usefulness of the AVA and warn against an overestimation of the AVA in the diagnosis of voice disorders. Moreover, phoniatricians strove continually to bring into line the AVA and the PVA. Many studies complete the AVA by the PVA and vice versa. So, the PVA of the pathological voice sound was developed in parallel to the AVA, but it was not paid so much attention as the AVA. Now, we have an increasing number of studies about the PVA. Because the evaluation of the pathological voice belongs to the medical domain, the methods of the PVA are directed to be practicable. The evaluated voice attributes are of medical (frequently of colloquial) nature and the evaluation scales show only a raw graduation. A typical example is the GRABS-scale [5], which did not fmd acceptance, at least in Europe. An excellent review about the PVA is given by KREIMAN *et al.* [9].

2. Problems with the Acoustical Analysis

1. The technology of the AVA is often used without sufficient knowledge of the matter. Among many others, two cases shall be mentioned here. (a) It is common not to pay attention to the measuring parameters. For instance, in the case of shimmer and jitter measurement when the sampling rate is too low. Then the amount of perturbation depends on the sampling rate (Fig. 1). (b) Methods from technical disciplines were applied to pathological voice signals without checking whether they are suitable, for instance linear predictive analysis. The method was originally developed to optimize speech coding in speech transmission.

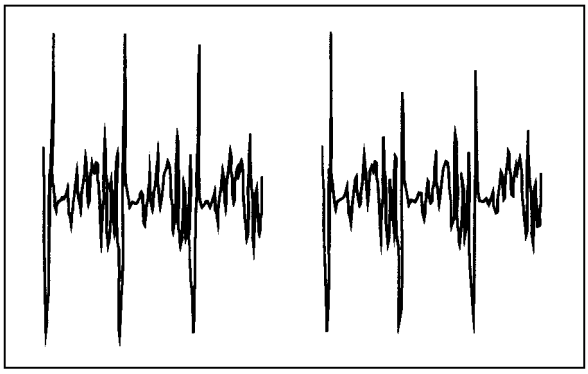

Fig. 1. Same voice periods, on the right sampled with the half sampling frequency of the left side.

2. Differences between different measuring systems (hard- and software, room acoustics, etc.) complicate a comparison of the measurement results [7].

3. Voice quality is affected apart from the influence of any disorder by age, emotion, vigilance, individual anatomy, language, dialect, individual speech behavior, test mate-

rial, etc. An extraction of the pathological features should be independent of all these influences, but this seems nearly impossible.

4. The voice signal may be characterized by a large number of different acoustic measurements. These characteristics may vary strongly in different situations (time of the day, emotional state etc.), and should be measured repeatedly. Maybe their number is limited in the case of sustained vowels. From running speech however many more characteristics could be extracted. This could be managed if there was an agreement about what to measure and about adequate standards [16]. This is however not in sight. By measuring many acoustical quantities DELIYSKI [4] proposed a way out whereby the quantities used for the evaluation is left to the investigator.

5. The measured quantities often show a direct or indirect correlation, and their true relationships are largely unknown. For instance, when prominent harmonics coincide with formant frequencies and the jitter shifts the harmonics within the formant peak, then the harmonics show fluctuations of their amplitude and hence shimmer arises (Fig. 2). Therefore, jitter creates shimmer, i.e. they correlate. The strength of the correlation depends on the vowel under consideration. In context of the measurement of shimmer, jitter measurement and formant analysis must always be performed to evaluate the shimmer magnitude. This is very seldom done.

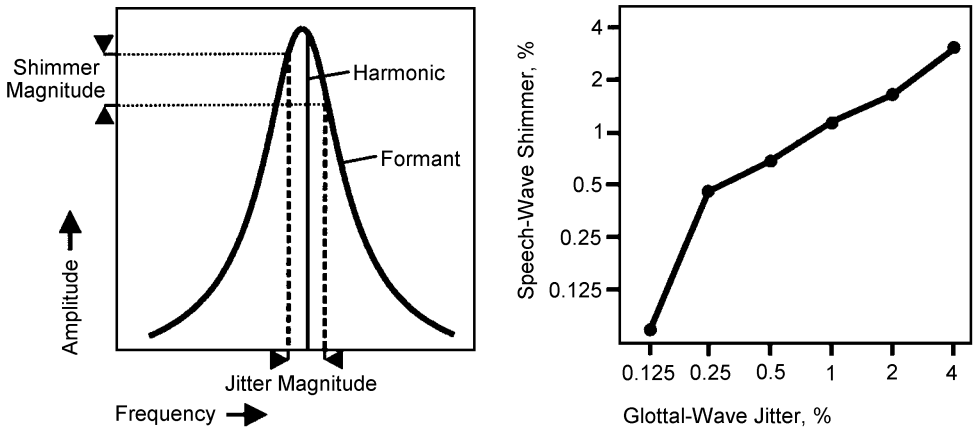

Fig. 2. Left: Illustration of the emergence of shimmer due to jitter. Right: Speech-wave shimmer due to glottal-wave jitter (source-filter model: fundamental frequency 125 Hz, vowel /a/)

6. The measured quantities of the voice signal do not show definite relationships to vocal physiology. For instance, noise in the voice signal could be created by a glottal chink as well as by any constriction in the vocal tract. Moreover, there are very different shapes of the glottal chink with different diseases. The noise however would hardly differentiate between these shapes.

7. BAKEN and ORLIKOFF [2] concluded that a diagnosis can hardly be performed by the AVA. They advanced several arguments which could be supplemented by the following;

(a) In spite of hundreds of papers about the AVA only a small number of studies used the AVA for real medical diagnosis of voice disorders.

(b) The transition between normal and pathological state is very wide and due to its multidimensional character can hardly be graduated.

(c) Optimized breath control and vocal technique may compensate for laryngeal pathology, making the voice signal less reliable as a measure of laryngeal health.

(d) Population means cannot be used as reference standards as voice production is highly individual.

(e) The diagnosis depends on a complex of symptoms and is of qualitative nature. On the contrary, the AVA yields only data of quantitative nature.

3. Problems with the Perceptual Analysis

1. The number of attributes for the description of voice features is very large and they are not standardized (aphonic, asthenic, breathy, coarse, creaky, diplophonic, dull, grating, guttal, heavy, hissing, hoarse, lax, light, metallic, nasal, poor, pressed, raucous, rasping, restrained, rough, rumbled, sharp, shrill, strained, strangled, tensed, thin, trembling, tremulous, unstable, veiled, waver, wet, wheezy, etc.). Moreover, the attributes do not exactly agree or compare with the functions of the vocal tract, the glottis, the manner of speaking, the speech accents etc.

2. Single listeners in contrast to homogenous listener groups differ in experience and individual perceptual habits and they show fatigue and mistakes in the experiments. Therefore, the evaluations are subjective.

3. Voices with extreme features are easy to classify, most voices are within the gray area between normal and severe. But just in this area the evaluation is unreliable, and the rating scores show lowest confidence (Fig. 3). Moreover, there are voices which cannot be reproducibly classified.

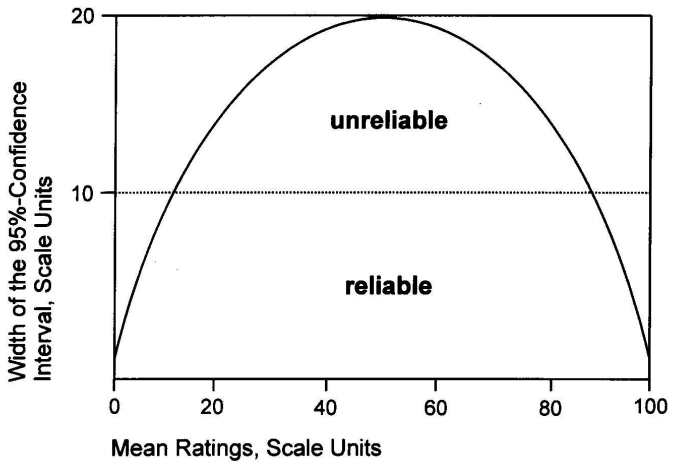

Fig. 3. Width of the confidence interval versus a rating scale (95% confidence interval = mean rating $\pm 1,96$ standard error of the mean), simplified for roughness ratings performed by KREIMAN *et al* [9].

4. The reliability and the agreement of the raters described in the literature [9] is insufficient for a passable classification of sound features of the pathological voice.

5. The evaluation procedure consumes much time.

4. Psychoacoustics

There have been a lot of attempts to integrate the AVA and the PVA. The effort did not succeed. This may be explained by a simple experiment. When the voice fundamental is systematically varied the region of the first formant, then the harmonics sample the formant along its contour from a minimum across a maximum to a minimum. The acoustical quantity sound pressure of the signals show a similar behavior as the formant contour, that is from a minimum across a maximum to a minimum. However, the psychoacoustical quantity loudness of the signal reveals hardly any variation. Here, the acoustics yields a result that is physically real but perceptually irrelevant, but the psychoacoustical quantity coincides with the perception.

Maybe psychoacoustics would yield a suitable and perhaps better methodology for the evaluation of the voice? Psychoacoustics would integrate the AVA and the PVA because the voice would be evaluated according to perceptive criteria and the objective character of the evaluation would be saved as psychoacoustics uses measurement methods.

There is an age old implication that psychoacoustics supports the physiological access to the problem of voice evaluation. When a subject doubles the subglottal pressure, the sound pressure increases by about 9 dB. This increase doubles the loudness sensation. Hence, doubling in the production is linked to doubling in the perception. This intra-human consistency becomes obvious only because a defined psychoacoustical quantity is used.

Psychoacoustical methods have to measure as we hear. Their problem is the reproduction of the sensation process by models. Such models are complex and difficult to construct.

In psychoacoustics we have measuring scales and procedures based on general perception rules resulting from listening experiments with respect to the characteristics of human hearing. One starts with a sound revealing a relevant attribute. Judges vary the magnitude of the attribute to match the criterion such as “half as” or “twice as”. Anchored sounds are used for a reference to get absolute magnitudes of the sound attributes. Then, a psychoacoustical attribute is like a physical quantity, it has a value and a dimension. An excellent introduction and review of psychophysics was written by Zwicker and Fastl [18]. In the following psychoacoustic categories are introduced and discussed with respect to the voice quality.

4.1. Loudness

Loudness is one of several base categories in psychoacoustics. It is our sensation due to sound-pressure amplitude measured on a scale between soft and loud and it is measured

in sone. Loudness can be computed from the spectral distribution of sound pressure [12] and can also be measured by a loudness meter.

4.2. Pitch

One of the important attributes of the voice sound is the pitch. When a subject is presented with the tuning standard of 440 Hz and then asked to choose the half-octave tone then a 220 Hz-tone is selected. However, when the procedure is repeated with a 8 kHz-tone the listener does not choose a 4 kHz-tone but a 1.3 kHz-tone instead. To adapt the human pitch sensation, the *mel*-scale is introduced. On the mel-scale, the half-octave tone of 2100 mel, that means 8 kHz, is half of 2100 namely 1050 mel, and that is 1.3 kHz.

The pitch of tones depend not only on frequency but also on the sound pressure. For instance a 200 Hz-tone of 80 dB produces a lower pitch than one of 40 dB. Additional sounds shift the pitch of a tone. Regarding these interactions the original acoustical pitch is transformed in a new pitch called spectral pitch. The transformation occurs in the inner ear. The actual perception, however, takes place on higher levels in our nervous system. Here, learned and stored patterns govern the perception. For a simple explanation we can think of a neural network, where the inputs are the spectral pitches of the voice overtones. Then these pitches are combined in a network by weighted connections between the single frequency tracks. These are learned connections in neuronal circuits. At the outputs of the network different pitches emerge, and the most prominent one we perceive as the pitch of the voice that differs more or less from the acoustically measured fundamental frequency [14].

An analogous example comes from the visual system. In Fig. 4 one can see the rectangle although it is not drawn. We have learned to recognise a rectangle from its contour. Even if the contour is not present we can see it. In pitch perception we hear the fundamental tone even when it is absent, for example in telephony. Here the overtones build the contour.

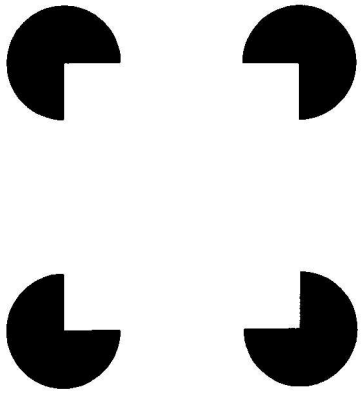

Fig. 4. Visible rectangle, which is not drawn.

In a similar way we also perceive the vowels since the overtone structures are the contours we have learned and stored. Maybe it can be assumed that we have stored patterns also for the voice qualities?

There are several pitches of a complex tone which are in competition. In the voice sound, these are pitches of the fundamental and the formants. The attribute pitch strength indicates the perceptive prominence of one pitch. So one can think of the measurement of tensed voice by means of the pitch strength. It is known that the voice fundamental dominates considerably with higher tension of the vocal folds or in breathy phonation because they vibrate in a more sinusoidal manner. On the other hand, this effect is also assigned to the psychoacoustical of categories tonality and sharpness.

4.3. Sharpness

A psychoacoustical item related to the voice sound is sharpness. Sharpness describes the sensation of spectral high-frequency components. Whether the components are tones or noise is of less importance. We hear the spectral components with different loudness along our perceptual frequency range. Therefore, sharpness can be computed by weighting the loudnesses in successive frequency regions, where higher regions are assigned higher weights. Then the weighted loudness parts are added up for the total sharpness. Sharpness is measured in *acum*. The reference sound of one acum is a small band noise at 1 kHz and 60 dB. Sharpness increases with increasing sound pressure, so by a factor of two from 30 to 90 dB. Moreover, because of the stronger growth of the number and energy of the higher harmonics with voice intensity, the sharpness is additionally increased.

4.4. Roughness

The central term used to described the acoustic impression of an organic voice disorder is hoarseness, a very undefmed term. The psychoacoustical sound attribute that meets this sensation is roughness. Roughness is a perceived sound quality that results from sound properties as well as the characteristics of the auditory system. Some qualities of the sounds and the sound processing system should be discussed first. The roughness eliciting sounds are signals with fluctuations of their envelope, such as amplitude or frequency modulated signals, pairs of beating tones, and pulse trains. With reference to voice, the amplitude modulation is either statistically distributed shimmer or a regular amplitude variation, for instance as in the case of bitonality, and jitter is a random frequency modulation.

The psychoacoustical concept of roughness is therefore responsible for different voice effects linked to organic dysphonia. Amplitude modulation below 20 Hz is perceived by its loudness fluctuation. The sound is verbally described by rattle or creak. Another psychoacoustical sound attribute – the fluctuation strength – describes this feature better than roughness. Fluctuations of higher frequencies are perceived as an unpleasant, disturbing component which is called roughness, raucousness, or harshness. In the upper frequency region above 300 Hz, the roughness sensation diminishes.

In the auditory evaluation of roughness we have two characteristics of the hearing system- a selective and an integrating feature. Selection means frequency resolution of the cochlea and integration means low pass behavior of the retrocochlear system. Between the sound field and the excitation stimulus of the neural system, there is a hydromechanical converter – the cochlea – which correlates a frequency region with a partition of the sensory area. These partitions are called critical bands. The auditory sensation differs when a sound event falls into one critical band or shares several bands. For example, a tone is only masked by noise when the noise is within the same critical band. The width of the critical bands determines the frequency resolution, above 500 Hz the width amounts to 20% of the frequency value, below 500 Hz the width is constant at 100 Hz (Fig. 5).

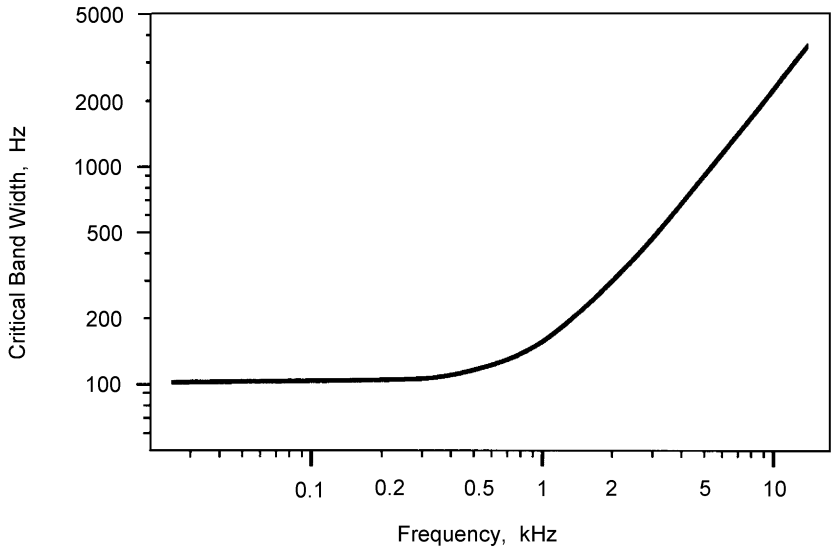

Fig. 5. Bandwidth of the critical bands in dependence on their mid frequencies.

For that reason we can only aurally resolve the first 6 to 10 harmonics of the voice. The integrative feature of the auditory system results from the neural transmission velocity which is limited to pulse rates of about 300 Hz. For the establishment of a measure an anchor sound is needed. In the case of roughness, this is an 100% amplitude-modulated 1 kHz-tone of 60 dB at a modulation frequency of 70 Hz. This tone has a roughness of one *asper*.

In the following some features of roughness are listed.

1. There are about 20 audible steps throughout the total range of roughness.
2. Roughness components from separate critical bands are added up to compute total roughness.
3. Roughness depends on loudness. An increment of sound pressure from 40 to 90 dB for an amplitude modulated tone doubles the roughness.

4. Roughness perception does not require exact periodical modulation. This fact is interesting with respect to jitter and shimmer.

5. The roughness reaches its maximum near modulation frequencies of 70 Hz.

6. The frequency modulation can produce much larger roughness than amplitude modulation. This is an important result. As can be seen from Fig. 6, 10% shimmer hardly modify the voice spectra, 10% jitter converts the harmonic structure into noise.

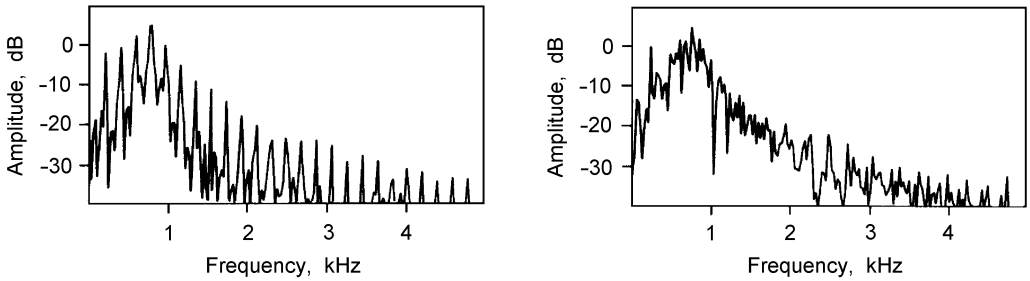

Fig. 6. Left: 10% shimmer, right: 10% jitter

7. Because of smaller damping in the higher critical bands, fluctuations pass these bands better and would create more roughness in the higher frequency region. However, the high partials in the voice are weak, and the roughness in this range is low. An exception is the singer's voice. Tenors and altos differ in the fourth and fifth formant. In tenors the formants are close together and fall within one critical band hence creating roughness, but in altos the timbre is changed because the formants are localized in different bands.

When roughness arises by amplitude modulation then it is obvious that roughness is an inherent feature of the non-pathologic voice because we have an alternation of high amplitudes at glottal closure and low amplitudes during the glottal open phase. Maximum roughness is expected at 70 Hz, therefore the male voice is more rough than the female voice. When one listens to the voice of Don Cossacks or Louis Armstrong then one can imagine how "rough" sounds. However, there is some confusion about normal and pathological roughness. A lot of work has to be done to clarify the problem. Perhaps we should differentiate between the more tonal roughness and the more irregular harshness, though pathological roughness can be created by regular variation due to the glottal amplitude modulation e.g. by a glottal neoplasm.

4.5. Fluctuation strength

Fluctuation strength is a sound attribute that arises in a very low frequency region. Its characteristics is of a band-pass type with a center frequency of 4 Hz. It is measured in *vacil*. Vocal vibrato and voice tremor represent high fluctuation strength.

4.6. Tonalness

The psychoacoustical categories coincide more or less with the voice qualities used in the PVA of pathological voices. The only perceptual feature without a psychoacoustical pendent is breathiness. It might be thought to measure breathiness by the attribute tonalness which describes how a sound differs from noise. The acoustical correlate of the psychoacoustical attribute tonalness is the ratio of harmonic to noise content in the signal (signal-to-noise ratio).

4.7. Sensory pleasantness

A more complex psychoacoustical attribute is the sensory pleasantness [1], called euphony in phoniatrics. It is influenced by elementary auditory sensations and is similar to an item of the semantic differential technique. Sharpness, roughness, tonality, and loudness affect the sensory pleasantness. In experiments, the sensory pleasantness has been determined (Tab. 1). Male voices show lower pleasantness than female voices due to their higher part of roughness. The reader will appreciate the attribute, as the human voice is more pleasant than a vacuum cleaner.

Table 1. Relative values of the sensory pleasantness for different signals, Terhardt and Stoll, [15].

Signal	Sensory Pleasantness
musical chord	1.0
female voice	0.77
male voice	0.65
vacuum cleaner	0.32

5. Application in phoniatrics

In literature, loudness is sometimes used instead of the sound pressure, as well as pitch instead of frequency. However, pitch-loudness patterns instead of the usual voice spectra are not found.

Imaizumi [6] has examined the roughness of disordered voices. In his discussion, he referenced his results to the roughness as defined by psychoacoustics. In a study by Burchardi [3], listeners with experience in psychoacoustical tests evaluated disordered voices. The results were highly reproducible and the psychoacoustical categories showed more usefulness than the features of the PVA. The study was only a first trial to introduce psychoacoustics in phoniatrics. The psychoacoustical models of human hearing sensations are very complicated. The construction of the models as well as their realisation in computer programs is still in progress, for instance [11]. In phoniatrics therefore, their introduction will be a matter for the future.

References

[1] W. Aures, *Berechnungsverfahren fuer den Wohlklang (die sensorische Konsonanz) beliebiger Schallsignale, ein Beitrag zur gehoerbezogenen Schallanalyse.* Dissertation, Technical University of Munich (1984)

[2] R.J. Baken and R.F. Orlikoff, *Voice measurements: is more better?*, Log. Phon. Vocol., **22**, 147–151 (1997).

[3] T. Burchardi, *Hoerversuche zur psychoakustischen Beurteilung gestoerter Stimmen*, MA Thesis, Technical University of Munich (1996).

[4] D.D. Deliyski, *Acoustic model and evaluation of pathological voice production* (*Multi-dimensional voice program model 4305*), Kay Elemetrics Corp. Issue A (1993).

[5] M. Hirano, *Clinical examination of voice*, Springer-Verlag, Wien New York 1981.

[6] S. Imaizumi, *Acoustic measures of pathological voice qualities - roughness-*, Ann. Bull. RILP, **19**, 179–90 (1985).

[7] M.P Karnell, R.S. Scherer and L.B., Fischer, *Comparison of acoustic voice perturbation measures among three independent voice laboratories*, J. Speech Hear. Res., **34**, 781–90 (1991).

[8] F. Klingholz, *Die Akustik der gestoerten Stimme*, Georg Thieme Verlag, Stuttgart New York 1986.

[9] J. Kreiman, B.R. Gerratt, G.B. Kempster, A. Erman and G.S. Berke, Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research, J. Speech Hear. Res., **36**, 21–40 (1993).

[10] *the Medical Equipment Jounal of Japan*, **26**, 1–6 (1982).

[11] PAK-System, *VibratoAkustikSystem*, Fa. Müller BBM, D-82152 Planegg-München

[12] E. Paulus and E. Zwicker, *Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln*, Acustica, **27**, 253–266 (1972).

[13] R.K. Potter, G.A. Kopp and H.C. Green, *Visible speech*, D. van Nostrand, New York 1947.

[14] E. Terhardt, *Zur Tonhoehenwahrnehmung von Klaengen*, Acustica, **26**, 173–199 (1972).

[15] E. Terhardt and O. Stoll, *Skalierung des Wohlklangs von 17 Umweltschallen und Untersuchung der beteiligten Hoerparameter*, Acustica, **48**, 247–253 (1981).

[16] I.R. Titze, *Toward standards in acoustic analysis of voice*, J. Voice, **8**, 1–7 (1994).

[17] H. Vogel, *Die Zungenpfeife als gekoppeltes System*, Ann. Phys., **62**, 247–282 (1920).

[18] E. Zwicker and H. Fastl, *Psychoacoustics – facts and models*, Springer-Verlag, Berlin Heidelberg New York 1990.