AURAL EXCITER AND LOUDNESS MAXIMIZER:

WHAT'S PSYCHOACOUSTIC ABOUT "PSYCHOACOUSTIC PROCESSORS" ?

Josef Chalupper

Institute for Human-Machine Communication, Technical University of Munich
80290 Munich, Germany
PH: +49-89-28928553 FAX: +49-89-28928535
e-mail: Josef.Chalupper@mmk.ei.tum.de

Abstract - In this study two so-called "psychoacoustic processors" are examined exemplarily by applying concepts, models and methods of scientific psychoacoustics. Physical measurements of processed sounds and results of hearing experiments on speech intelligibility and sound quality (Aural Exciter) and loudness (Loudness Maximizer) are presented and discussed with regard to classic psychoacoustic models and potential new applications. Therefore, relevant psychoacoustic facts, in particular on perception of nonlinear distortion, are reviewed.

I.   INTRODUCTION

Nowadays many so-called "psychoacoustic processors" are commercially available, but independent scientific investigations of these devices are very rare. Moreover, psychoacoustic hearing sensations which those processors are said to influence, and psychoacoustic phenomena on which their functional principle is based, are often described with unclear, self-invented "psychoacoustic" terms.

It even appears that "psychoacoustic processors" are deliberately surrounded by mystique to increase their appeal. In contrast, this investigation uses scientific methods to find out what is psychoacoustic about those devices. Since both Steinberg's Loudness Maximizer and Aphex' Aural Exciter are commonly used tools in mastering and broadcasting, they are examined exemplarily in this study.

Psychoacoustics is a branch of acoustic science, which - in contrast to electroacoustics - investigates sound not only from the physical, but also from the human - or psychological - point of view. More specifically, the task of psychoacoustics is to develop functional models, which relate physical parameters of an acoustic stimulus to hearing sensations evoked in human listeners. Since the human auditory system is the final receiver in almost all cases of sound recording, transmission and reproduction, its properties should be taken into account in all fields of audio engineering [1].

Section II will briefly review some basic psychoacoustic facts and models, which are important for understanding this study. Further information about psychoacoustics is available from [2].

In order to be able to answer the title question, it first seems necessary to come to terms with a definition of "psychoacoustic processors". Therefore, throughout this study psychoacoustic processors are defined as audio signal processors that fulfill at least one of two criteria:

(1) There must be a measurable difference between processed and unprocessed sounds, concerning a specific hearing sensation, while all other hearing sensations are nearly unaffected.

(2) The functional principle takes into account psychoacoustic knowledge, e.g. masking effects, auditory time resolution etc.

If, for instance, loudness is raised without any perceptual difference in fluctuation strength and sharpness, criterion 1 would be fulfilled. On the other hand, a MPEG-codec is "psychoacoustic" in terms of criterion 2, because its algorithm uses masking effects.

In section III results of physical measurements and hearing experiments are presented to check what - if at all - criterion is fulfilled by the Aural Exciter and the Loudness Maximizer, respectively.

II.   FUNDAMENTALS

II.1   Relevant Psychoacoustic Facts & Models

Masking & critical bands
A very basic concept in psychoacoustics is the assumption that the human auditory system analyses incoming sound like a bank of overlapping filters. These filters are called critical bands and have below 500 Hz a constant absolute bandwidth of about 100 Hz while above 500 Hz they have a constant relative bandwidth of about a third octave.
Hence, frequency can be transformed to a hearing equivalent scale, that is, the critical band-rate scale z (or 'Tonheit'). The unit of this scale is called 'Bark', which corresponds to the bandwidth of one critical band [2].
Closely related to the critical band concept is the effect of masking. Masking takes place both in the time ('nonsimultaneous masking') and the frequency domain ('simultaneous masking'). Fig. 1 shows (simultaneous) masking patterns of a narrowband noise for different levels. Note that the slope towards high frequencies gets shallower with increasing masker level.
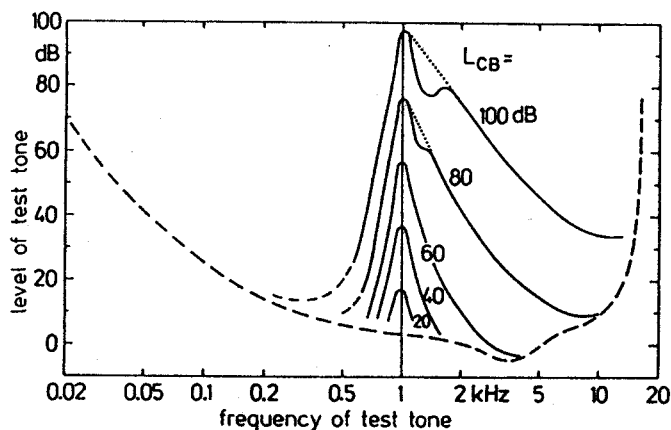

Fig. 1. Level of test tone just masked by critical band wide noise with centre frequency of 1 kHz and different levels as a function of the frequency of the test tone (adopted from [2])

Formulas for calculating masking patterns on the basis of a signal's spectrum are given by Terhardt [3]. The slope S_1 of a single spectral component's masking pattern towards lower z values is

$$S_1 = 27 \text{ dB/Bark} \tag{1}$$

while the slope towards higher z values, S_2, depends on level and frequency:

$$S_2 = [24 + 0.23(f_M/kHz)^{-1} - 0.2L_M/dB] \text{ dB/Bark}, \tag{2}$$

where f_M and L_M denote the masker's frequency and level, respectively.

Recently a computer program for calculating nonsimultaneous masking has also been published [4].
A practical application of these effects is perceptual audio coding, since irrelevant information can be reduced without introducing distortions by taking into account masking patterns [5].

Loudness
As can be seen from figure 2, the loudness of sounds with the same level but different spectra can vary markedly. For a loudness of 1 sone, a 1 kHz sinusoid must have 40 dB SPL, whereas a broad band noise reaches the same loudness at about 30 dB. For levels above 40 dB, a doubling of loudness is achieved with each increment of 10 dB.
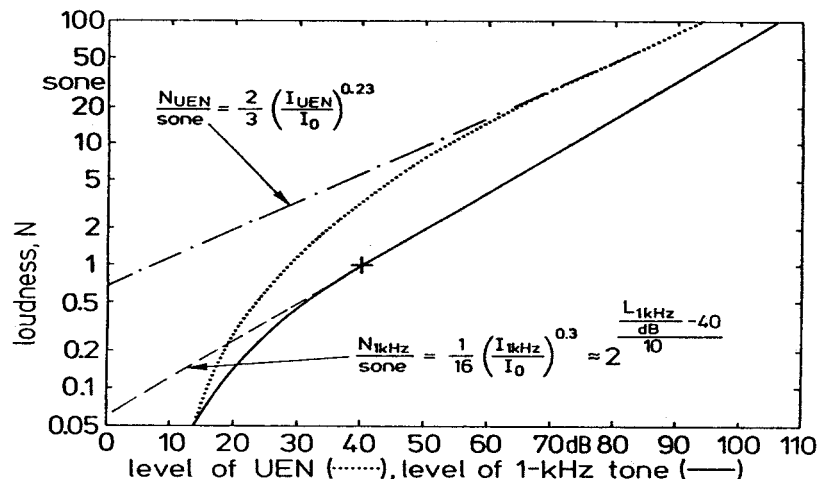
Fig. 2. Loudness function of a 1-kHz tone (solid) and of Uniform Exciting Noise (dotted). Approximations using power laws are indicated as broken and as dashed-dotted lines together with their corresponding equations (adopted from [2])

From psychoacoustically measured loudness of Uniform Exciting Noise, a function relating loudness and level in a single critical band can be deduced [2]:

$$N'(z) = N_0 \left(\frac{E_{THQ}(z)}{s(z)E_0} \right)^{0,23} \left[(1 - s(z) + s(z) \frac{E(z)}{E_{THQ}(z)})^{0,23} - 1 \right], \qquad (3)$$

where N' is the specific loudness in sone/Bark, E excitation (corresponds to the level in one critical band) and E_{THQ} excitation at hearing threshold. Total loudness is obtained by integrating specific loudness across all critical bands.
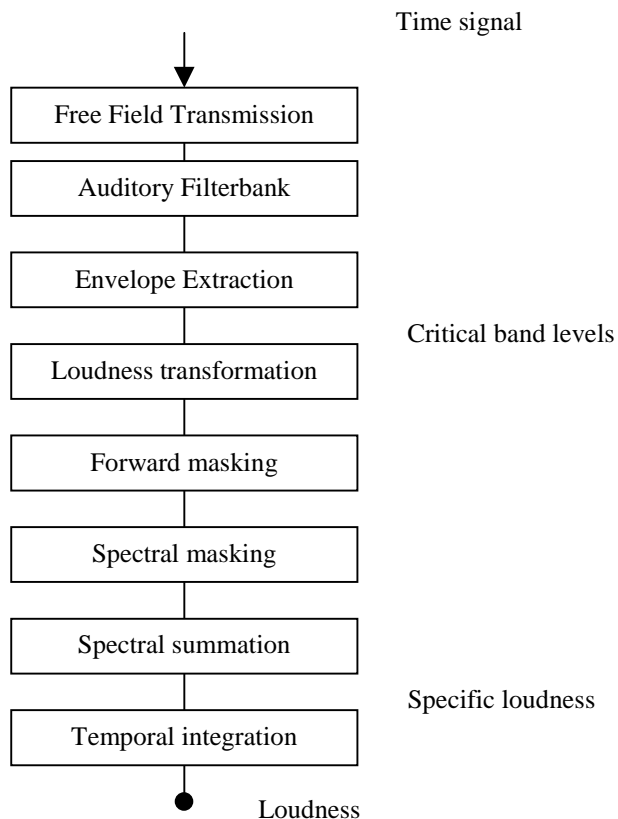
Time signal

Fig. 3. Block diagram of the dynamic loudness model

In order to calculate also loudness for time varying sounds, dynamical effects like forward masking and temporal loudness integration [2] have to be considered. The block diagram of a recent implementation of the dynamic loudness model [6] is depicted in figure 3.

The first stage incorporates transmission from free field to the inner ear by a fixed filter. After analysing incoming sound with an auditory filter bank, in the box 'envelope extraction' short term RMS levels are calculated within an aurally adequate temporal window. The filter bank is implemented by a fourth order Fourier Time Transformation [7], [8], with 24 analysis frequencies spaced by one Bark. The equivalent rectangular bandwidth of the resulting analysis filters is set to 1 Bark. The temporal window is chosen according to [9], with a duration of 8 ms. Critical band levels then are transformed into specific loudness by eq. (3). By taking into account masking effects ('forward masking', 'spectral masking'), one gets the specific loudness time pattern, which is regarded as an aurally adequate representation of sound. As will be shown later, specific loudness time pattern is a prerequiste for understanding the Aural Exciter and the Loudness Maximizer. After spectral and temporal integration, time varying loudness is obtained. If loudness fluctuates markedly as a function of time, perceived 'global loudness' corresponds to N_5, which is the loudness that is exceeded in 5% of the time. The stationary part of this model is standardized

in DIN 45631 [10]. This loudness model can be easily fitted to individual hearing losses [6]. A simplified version of the dynamic loudness model is used for predicting speech quality [11].

Sharpness

A factorial investigation on verbal attributes of timbres of steady sounds has shown that the attribute sharpness reperesents the factor carrying most of the variance (44%) [12], and thus seems to be more suitable for the description of timbre than other scales like density [13]. It was found that the sharpness of narrow band noises increases proportionally with the critical band rate for center frequencies below about 3 kHz. At higher frequencies, however, sharpness increases more strongly, an effect that has to be taken into account when the sharpness S is calculated using a formula that gives the weighted first momentum of the specific loudness pattern:

$$S = 0.11 \frac{\int_0^{24\,Bark} N'\cdot g(z)z\,dz}{\int_0^{24\,Bark} N'\,dz}\, acum \, . \tag{4}$$

In equation (4), the denominator gives the total loudness, while the upper integral is the weighted momentum mentioned. The weighting factor $g(z)$ takes into account the fact that spectral components above 3 kHz contribute more to sharpness than components below that frequency [2].

Fluctuation strength & roughness

These hearing sensations are correlated to the temporal variations of sounds. Fluctuation strength measured as a function of modulation frequency shows a maximum near 4 Hz, whereas roughness can be described by band pass characteristic at 70 Hz. This means that very slow variations (< 0.5 Hz) hardly affect these dynamic hearing sensations. Another important fact is that roughness and fluctuation strength increase with increasing modulation depth up to about 30 dB, where a saturation can be observed. Both roughness and fluctuation strength can be calculated from the specific loudness time pattern [2].

The above mentioned hearing sensations and their models are applied very successfully in sound quality design [14]. How much loudness, sharpness, roughness and fluctuation strength a specific sound needs, however, can not be answered generally, since this depends strongly on the sound itself and the environment, where it will be used. For example some roughness can give the sound of a sporting car the right flavour, but may spoil the sound quality of a family van.

II.2.   Perception of Nonlinear Distortion

Since nonlinearities play a major role in both devices, it is important to know how distortions are perceived by human listeners. The discussion will be in two parts:
The first on the physical description of nonlinear distortions, the second on the perception of nonlinear distortions.

From a physical point of view, a memoryless nonlinearity can be modeled as a polynomial of the form:

$$y = a_1 x + a_2 x^2 + a_3 x^3 + \ldots\ldots + a_n x^n \tag{5}$$

where x is the input and y is the output of the nonlinearity. Since derivations even for simple input signals are both longwinded and tedious, we will confine to quadratic and cubic distortions of a single sinusoid.
For a quadratic nonlinearity where the transfer characteristic is

$$y = a_2 x^2 \tag{6}$$

the distortion from an input signal

$$x = A\sin(\omega t) \hspace{8cm} (7)$$

can be simply obtained by the use of trigonometric relationships:

$$y = a_2 x^2 = \frac{1}{2} a_2 A^2 - \frac{1}{2} a_2 A^2 \cos(2\omega t) \hspace{5cm} (8)$$

In the case of a cubic distortion

$$y = a_3 x^3 \hspace{10cm} (9)$$

one gets for the same input signal:

$$y = a_3 x^3 = \frac{3}{4} a_3 A^3 \sin(\omega t) - \frac{1}{4} a_3 A^3 \sin(3\omega t) \hspace{4cm} (10)$$

From (8) and (10), it can be seen that quadratic nonlinearities lead to distortions at twice the signal frequency, whereas cubic nonlinearites produce distortions at 3ω, which are the second and the third harmonic, respectively. For n order nonlinearities it can be stated that odd order nonlinearities produce odd harmonics (1, 3, 5 ...), and even order nonlinearities even harmonics (2, 4, 6 ...) between the signal frequency ω and nω. If the input signal contains more than one sinusiod, distortions at lower frequencies are also introduced; for example, two sinusoids with frequencies ω₁ and ω₂ produce a difference tone at ω₁ - ω₂ [15].
If the amplitude A of the input signal in (8) and (10) is raised by 6 dB, quadratic distortion increases by 12 dB and cubic distortion by 18 dB. This means that the amplitudes of distortions are heavily dependent on the level of the input signal.
In general the order of a polynomial for approximating a nonlinearity increases if its shape is very 'edged' [16]. Therefore smoothed curves ('soft knee') produce less distortions especially at higher frequencies. This is of great importance for digital implementations of nonlinearities, because high order distortions may exceed the nyquist frequency and due to aliasing show up at unexpected frequencies.

Now that we know about the physical aspects of nonlinear distortions, the question arises how human listeners perceive these distortions.
Between 1950 and 1960 listening tests on the detection of distortions were carried at the Technical University in Stuttgart, especially by Gäßler [17]. Based on the results a theory of perception of nonlinear distortion was developed for simple, stationary sounds, which can be qualitatively stated as follows:

If one or more of the distortion products are above threshold (hearing or masking threshold), the distortion is perceptible.

Therefore, given a signal and a nonlinearity, it is possible to determine whether distortions are detectable by calculating the signal's masking pattern from eq. (1) and (2) and the levels of all distortion products from (5). In practical applications a nonlinearity can be suited to a signal without perceptible distortions. If the signal is unknown, it is sufficient to analyse its (short-time-)spectrum. Thus, based on a Short-Time-Fourier-Transform, an adaptive algorithm can be designed to find a polynomial that approximates a desired nonlinearity without introducing audible distortions.

In 1982, Günthersen [18] extended the theory of perception of nonlinear distortion to complex, non-stationary sounds. He concluded that two different mechanisms can determine the threshold for detecting distortions:

1. *Direct perception of distortion products*
2. *Fusing of signal and distortion products to form a new percept.*

For simple, stationary signals the first of the mechanisms always determines the threshold, while for non-stationary signals threshold is always determined by the second mechanism. For complex, stationary signals both mechanisms can determine the threshold depending on peculiarities of the signal.

In 'Gestalt' psychology, it is generally assumed that multiple single objects can under certain conditions ('Gestalt laws') fuse to one single object. If those conditions - for example coherence - are not fulfilled, they will seggregate. A set of time-variant sinusoids, for example, can fuse to a single 'auditory stream' [19], if they are modulated coherently.

Since real-world signals, like speech and music are usually complex and non-stationary, at first sight, it seems as if the second mechanism is the most important one. Otherwise, distortion products that are completely masked by the original signal will not be able to form a new percept. Therefore, Gäßler's theory of perception of nonlinear distortions is also true for complex, nonstationary sounds; but in contrast to stationary sounds, from exceeding masking threshold follows not necessarily that sounds are perceived as distorted.

The additional harmonics change the spectral shape of complex sounds and thus, their sharpness.

Whether this variation in sharpness is perceived as being pleasant or unpleasant, depends strongly on the signal and therefore, it is possible that human listeners under certain circumstances even prefer the new - 'distorted' - signal.

Thus, listening tests have to be carried out to assess the influence of supra-threshold distortions for a complex, non-stationary signals.

In conclusion, we can state that the perception of nonlinear distortion is determined by masked threshold and fusing.

Masked threshold can be calculated and therefore easily used in an adaptive algorithm for controling nonlinearities without introducing audible distortions.

Supra-threshold distortions can fuse to a new percept, which might be - depending on the signal - preferable to the undistorted signal.

III.   Investigations on "psychoacoustic" signal processing devices

The methodology used to check the criteria mentioned in the introduction was essentially the same for both the Aural Exciter and the Loudness Maximizer:

Firstly, psychoacoustic experiments were carried out to assess how certain hearing sensations are affected by those devices. To exclude binaural effects, sounds were presented monaurally or diotically throughout this study.

Secondly, based on physical measurements, simple block diagrams were developed to explain functional principles of both psychoacoustic processors. Software implementations of these block diagrams achieve nearly the same perceptual effects as the originals, although they only share the principle, but differ markedly in detail.

Thirdly, psychoacoustic facts and models as described in the foregoing section are used to relate results of psychoacoustic and physical experiments and to answer the title question.

Finally, further applications are discussed.

III.1.   Aural Exciter

According to Aphex, the Aural Exciter will recreate and restore missing harmonics; when added, they restore natural brightness, clarity and presence, and can actually extend audio bandwidth. There are also some speculations in non-scientific literature about a speech enhancement effect caused by the Aural Exciter [20]. The only scientific study concerning speech intelligibility, which the author is aware of, was done by Herberhold [21]. He found a small, but statistically significant increase in intelligibility for speech in quiet and monaural listening. No significant improvement was found for speech in noise. All measurements were carried out with hearing impaired listeners equipped with hearing aids. The ambiguity of Herberhold's results may be caused by examining patients with differing degress of hearing loss and different hearing aids.

Thus, measurements of speech intelligibilty with normal hearing listeners are presented in this study. In addition, sound quality was assessed, since the original purpose of the Aural Exciter was to improve sound quality of music recordings.

III.1.1. Psychoacoustic Measurements

Speech Intelligibility

Speech intelligibility was measured in different noises with a German monosyllabic rhyme test ('Sotscheck-Test' [22]) and a German sentence test ('Marburger Satztest') [23]. For calibrating the Sotscheck-Test, a speech shaped noise according to CCITT Rec. G.227 ('CCITT-noise') was used, whose level equals the median of the L_{AFmax} of all (900) words. L_{AFmax} denotes the maximum A-weighted and with time constant 'fast' measured level of a single word. In the case of the 'Marburger Satztest' the calibration signal from CD [23] was taken.

Besides the stationary CCITT noise, a Harmonic Complex Tone ('HCT') with the spectral envelope of CCITT-noise and $f_0 = 100$ Hz, and a fluctuating noise proposed by Fastl ('Fastl-noise' [24]) served as interfering noises, which were always presented at a level of 65 dB SPL.

Speech material and noise were amplified and added to obtain the desired signal-to-noise ratio (SNR) and then fed into the Aural Exciter. Its output signal was free-field-equalized [2] and presented monaurally over headphones (Beyer DT 48).

Parameters of the Aphex Aural Exciter Type III (Model 250) generally were set according to manufacturer's recommendations for improving AM-radio [25], except if stated otherwise.

Eight normal hearing listeners took part in experiments 1 - 4, where the 'Sotscheck-Test' was applied (1 test list per subject), whereas 11 normal hearing listeners were used in experiment 5 for the 'Marburger Satztest' (2 test lists per subject).

In experiment 1, a slightly reverberated - for details see experiment 3 - Sotscheck-Test was presented in Fastl-noise at two different signal-to-noise ratios. Figure 4 shows that speech intelligibility can indeed be improved by an Aural Exciter for a signal-to-noise ratio of -10 dB. This improvement is statistically significant (Wilcoxon-Test: p=1.17%) and amounts to 10.8%. In contrast, at -15 dB (SNR) there is only a very small increase of speech intelligibility, which is not significant.
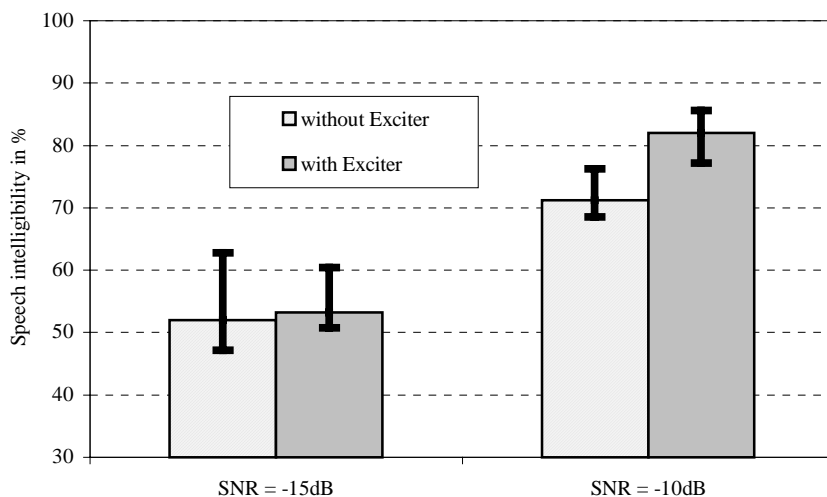

Fig. 4. Speech intelligibility in Fastl-noise with and without Aural Exciter at different signal-to-noise ratios.

Obviously the Aural Exciter does not act as speech enhancer at low signal-to-noise ratios. Thus, in the following experiments signal-to-noise ratio was chosen so as to ensure that speech intelligibility without exciter is about 70%.

To check whether this speech enhancement is due to the Exciter's linear or nonlinear distortions, in experiment 2 nonlinear distortions were excluded as far as possible by turning the 'harmonics' knob to 'min'. In so doing, speech intelligibility - see figure 5 -increases only by 2.4% (Wilcoxon-Test: p=6.25%). It should be noted that the reference measurement without Exciter in experiment 1 resulted in a somewhat lower (4%) speech intelligibility compared

with experiment 2, although stimuli were physically identical. Compared to the reference condition in experiment 2 the Exciter's linear distortions lead to a significant increase of 6.4 %.
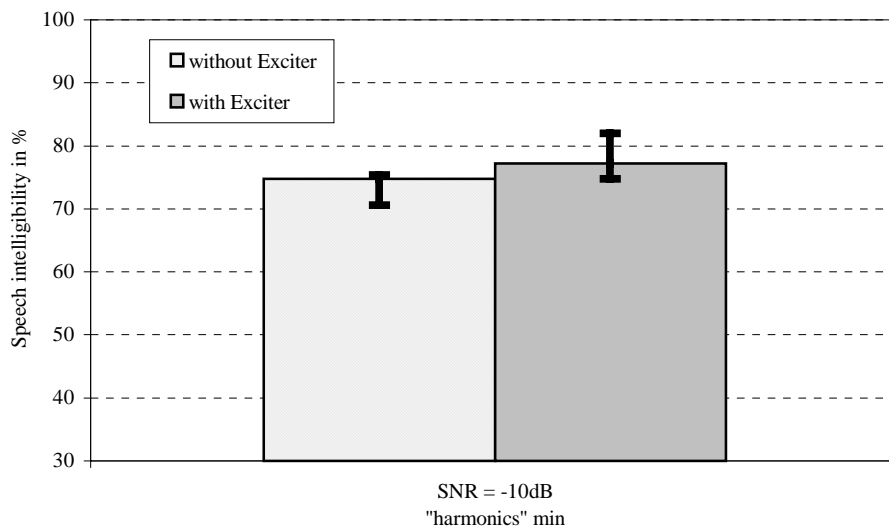

Fig. 5. Speech intelligibility in Fastl-noise with and without Aural Exciter for 'harmonics' set to min.

Summing up we can state, that the increase in speech intelligibility is at least partly due to nonlinear distortions. Since interquartile ranges span over about 10% and the amount of speech enhancement is close to 10%, it is difficult to obtain statistically significant results.

Next, the influence of reverberation is considered. In experiment 1 and 2, a slight reverberation (reverberation time =1s, reverberation level = -6dB) was added to speech and noise. The results of experiment 1 are replotted in figure 6 and compared to (new) measurements with strong reverberation (reverberation time =4s, reverberation level = 0dB) and without reverberation, respectively. With increasing reverberation a higher signal-to-noise ratio was chosen to obtain nearly the same intelligibility in the reference condition. While there is a small, but not significant increase in speech intelligibility without reverberation, no improvement was found in the case of strong reverberation. The latter might be due to the chosen signal-to-noise ratio, since the speech intelligibility of the reference condition is only 60% (see experiment 1).
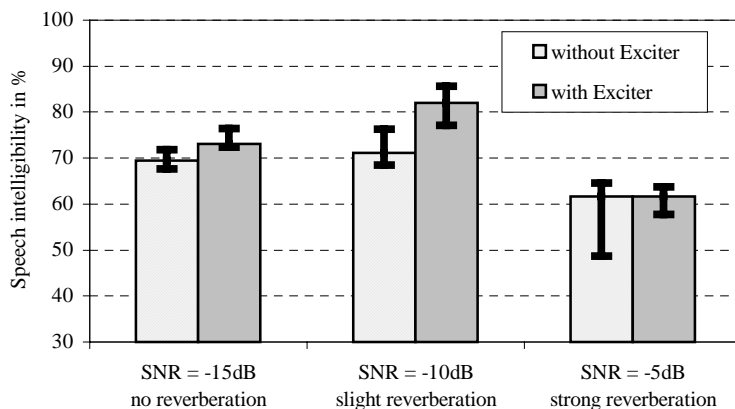

Fig. 6. Speech intelligibility in Fastl-noise with and without Aural Exciter for different reverberation times.

Thus, it is not possible to draw final conclusions concerning the influence of the Exciter on speech.

Figure 7 shows the results of experiment 4, where instead of the fluctuating Fastl-noise a stationary CCITT-noise was used. Speech intelligibility increases by 7.2%, but due to the large variance in the reference measurement (without Exciter) this is not significant. Again, signal-to-noise ratio might have been too low.
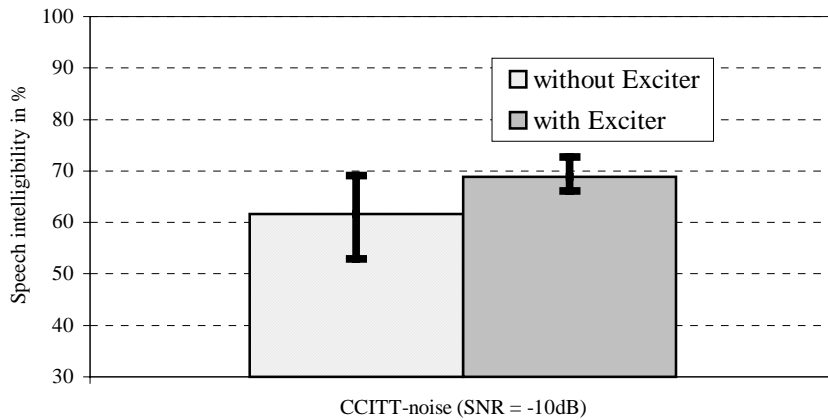

Fig. 7. Speech intelligibility in CCITT-noise with and without Aural Exciter

In general, there seems to be a trend towards enhanced speech intelligibility if speech is processed by an Aural Exciter. In certain cases intelligibility can be increased by about 10%, which is difficult to measure with statistical significance due to the limited accuracy of speech intelligibility measurements.

Therefore, in experiment 5 a sentence test was used. Typically sentence tests show a steeper discrimination function than monosyllabic speech tests. Moreover, the intelligibility of sentences is regarded as being 'more natural', since usually people talk to each other by means of sentences. Sentence intelligibility is defined as the percentage of correct words per sentence. Two noises were used: Fastl-noise at -8 dB signal-to-noise ratio and a HCT at -11 dB. Again, the results as depicted in figure 8 are ambiguous. While intelligibility increases in both cases, statistical significance is reached only for HCT. Moreover, the difference between excited and non-excited speech is 8% for Fastl-noise and 18% for HCT. Interestingly, this large and significant increase in speech intelligibility was obtained, although speech intelligibility at the reference condition is rather small (48%).
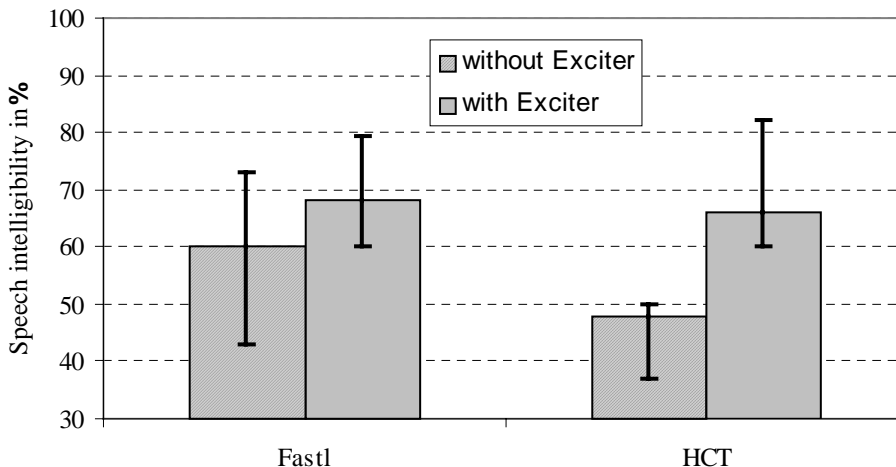

Fig. 8. Sentence intelligibility with and without Aural Exciter

The outcome of these experiments may be summarized as follows:

1. In special cases speech intelligibility can be increased by the Aural Exciter.
2. Increased intelligibility is less than 11% for monosyllables and less than 18% for sentences.
3. Since the amount of improvement is in the range of measurement accuracy, it is difficult to obtain statistical significant results.
4. Both linear and nonlinear distortions seem to contribute to speech enhancement.

Sound Quality

Sound quality was assessed for a great variety of sounds, namely
- timetable announcement at a railway station ('announcement')
- conversation in a restaurant ('restaurant')
- church organ ('organ')
- conversation at office ('office').

All sounds were recorded digitally on DAT with a sampling rate of 44.1 kHz, equalized to the same RMS level and low pass filtered (cut off frequency: 4kHz), which is typical for today's hearing aids. In the experiment sounds were presented diotically to exclude binaural effects at a level of 65 dB SPL with a free-field-equalized DT48.
Parameters of the Aural Exciter again were set according to manufacturer's recommendations for improving AM-radio [25].
Subjects had to listen to two pairs ('A' and 'B') of stimuli. Only one out of the four stimuli is aurally excited and its position in the trial is chosen at random. The task was first to indicate the pair containing a sound different from the other three. Then subjects had to judge the difference in quality by means of a scale between +5 and -5, corresponding to 'very much better' and 'very much worse'. This judgement was only valid, if the correct pair was indicated before. Subjects were encouraged to describe briefly in what way sounds differed.
Each sound was presented four times to eight listeners. The correct pair was identified in 91 % of all judgements, which indicates that the Exciter has a distinct effect on sound quality. As can be seen from figure 9, this effect depends strongly on the stimulus.
A clear improvement of sound quality was found for 'organ' and for 'office', whereas for 'restaurant' sound quality only increases slightly and for 'announcement' even decreases.
Taking into account interquartile ranges, we can state that the Exciter works best for musical signals and has no detrimental effects on speech sounds, but in special cases sound quality may decrease.
Processed sounds often were decribed as being 'sharper' and 'more brilliant' and in the case of 'announcement' as 'distorted'. Two subjects reported that the loudness of speech relative to the background noise was increased especially for 'office'.
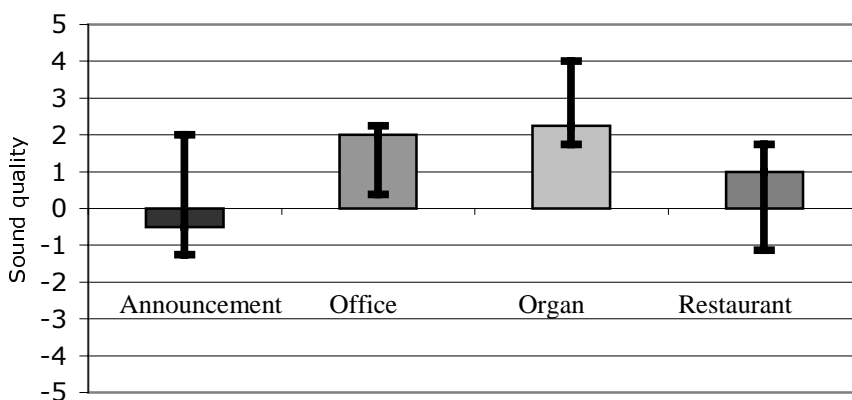

Fig. 9. Sound quality evaluation of 'aurally excited' sounds

To enable a discussion of these results in detail, it is necessary to understand the physical properties of the Aural Exciter.

III.1.2.   Physical Measurements

Instead of listing hundreds of data in detail, this section seeks to concentrate on the basic principles of how signals are processed physically by the Exciter.

Figure 10 shows the frequency response for the same parameter setting as used in the listening tests. Above about 700 Hz amplification becomes evident, which saturates at 7 dB for frequencies higher than 3000 Hz.
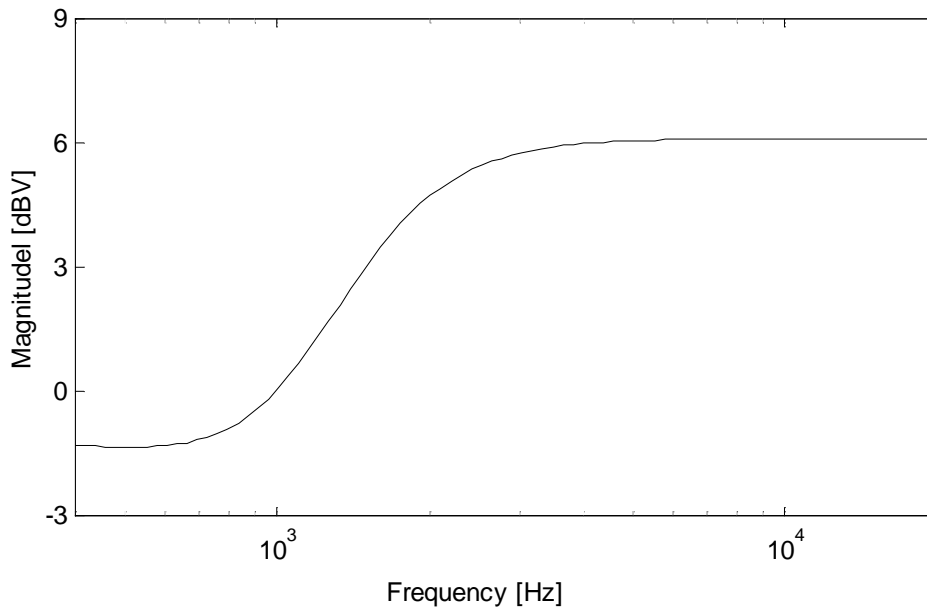

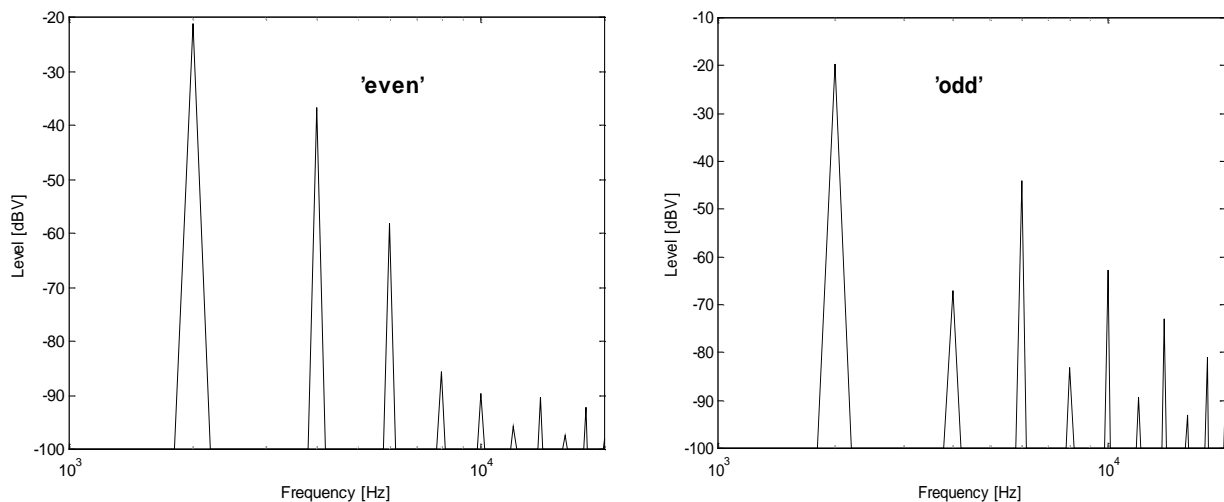**Fig. 10.** Frequency response of the Aural Exciter

Fig. 11. Nonlinear distortions for 'odd' and 'even' setting of 'timbre' for a 2 kHz tone input

Besides linear distortions, nonlinear distortions can also be measured. According to the manual, the Aural Exciter is capable of producing odd and even harmonics by adjusting 'timbre' to 'odd' or 'even', respectively. This is verified in figure 11, where the spectra of the output signals are depicted, when a 2 kHz tone is used as input. In position 'even' there are only even harmonics, whereas in position 'odd' odd harmonics dominate, but even harmonics still are measurable.

These nonlinear distortions combine with the original signal, whereby spectral shape is changed. Thus, with a third octave equalizer (Klark Teknik DN 27A), the Exciter's output was equalized such that if white noise is used as input, a flat spectrum is obtained at the EQ's output. If 'harmonics' is turned to 'min', frequencies above 2 kHz are attenuated by nearly 2.5 dB, whereas in position 'max' a amplification of 0.5 dB is measured.

The presented measurements reveal the essential characteristics of the Aural Exciter:
1. Low frequencies are unaffected.
2. Linear distortion boosts high frequencies by about 7 dB.
3. Nonlinear distortions can add an extra 3 dB increase and extend bandwidth (see section II.2).

The circuitry of Aphex' Aural Exciter is described in detail in the service manual [25]. A rather simple block diagram that can explain the basic principle of the Exciter is given in figure 12.
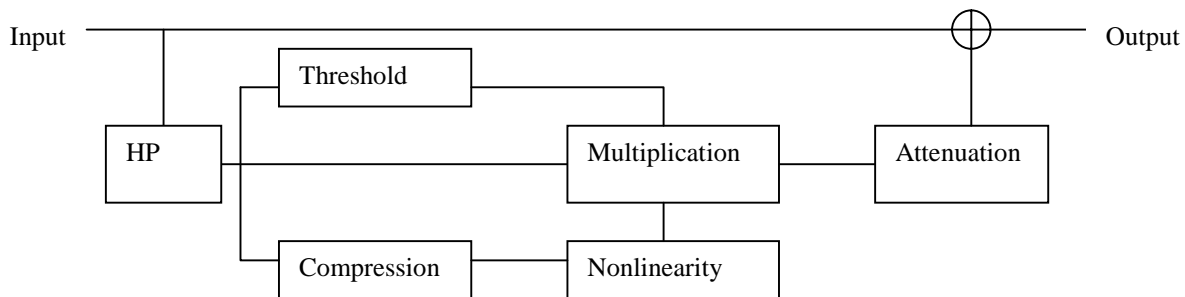For further experiments and investigations, this block diagram also was implemented in Matlab.

Fig. 12. Block diagram of the 'TUM Exciter'

The input signal is split into a main path and a sidechain. All processing is done in the sidechain, where the signal first is high pass filtered.
Its cut off frequency allows the frequency range to be contolled, where linear and nonlinear distortions become effective. For speech enhancement $f_c = 1000$ Hz seems appropriate. Details of implementation like filter order, phase response and group delay only play a minor role, which was verified in informal listening tests.
Threshold is controlled by a downward expansion. If input level is below threshold, distortions introduced by the Exciter are reduced. The input-output function of this downward expansion is given in [25].
The block 'compression' ensures that the following nonlinearity is fed by a nearly constant voltage. This is important to achieve similar distortion spectra for different input levels, since nonlinear distortion usually is very level dependent (section II.2).
The 'nonlinearity' can be any nonlinear function. Aphex' Aural Exciter uses halfway and fullway rectification to produce odd and even harmonics, respectively. Our implementation ('TUM Exciter') works with a polynomial to give a maximum of flexibility. A desired distortion spectrum is realized by varying polynomial coefficients (section II.2).
The outputs of 'threshold', 'nonlinearity' and 'HP-Filter' are multiplied, attenuated and added to the main path. Note that multiplication in the time domain corresponds to convolution in the frequency domain. Therefore, to produce linear distortion a DC-component is necessary.
If parameters are set appropriate, the 'TUM Exciter' acts physically identical to Aphex' Exciter. To verify, that the 'TUM Exciter' achieves the same perceptual effects as the original, experiment 4 (speech intelligibility in CCITT-noise) was repeated and compared to the results obtained with Aphex' Exciter. The TUM Exciter increases speech intelligibility by nearly the same amount, which also is statistically not significant. Even if onset consonants, vowels and final consonants are analyzed separately, results are still very similar. In particular onset consonants and vowels benefit.
Obviously it is sufficient to take into account linear and nonlinear distortions to understand the Aural Exciter's effects on speech intelligibility. Informal listening tests suggest that the same is true for sound quality.
Lindblad [26] investigated the influence of distortions on speech intelligibility and concluded, that distortions are detrimental especially to vowels, since new formants are created by fusing. Because this contradicts our results, additional physical measurements on synthesized vowels were performed.

Figure 13.1-3 shows spectra of a synthesized German /e/. In figue 13.1 the original spectrum is depicted , in 13.2 the spectrum of the same vowel, but after processing with the TUM Exciter and in 13.3 the same as in 13.2, but without applying a high pass in the Exciter's sidechain. While in figure 13.2 the level of the 2. and 4. formant is increased relative to f_0, in figure 13.3. a new formant is evident around 1200 Hz. The former is probable to result in an increase, and the latter in a decrease of speech intelligibility. Since Lindblad did not use a high pass filter, degraded intelligibility of vowels is obtained.
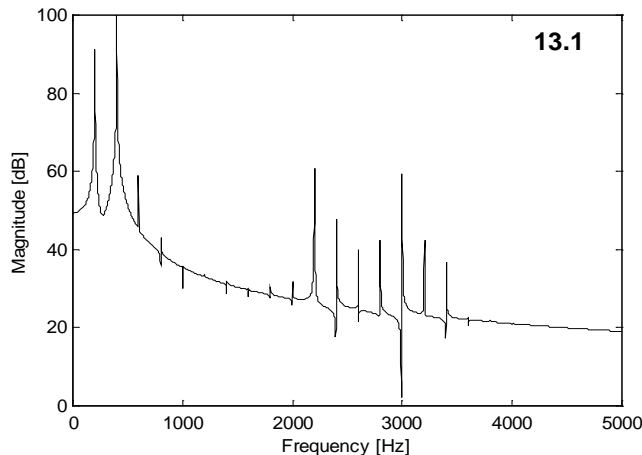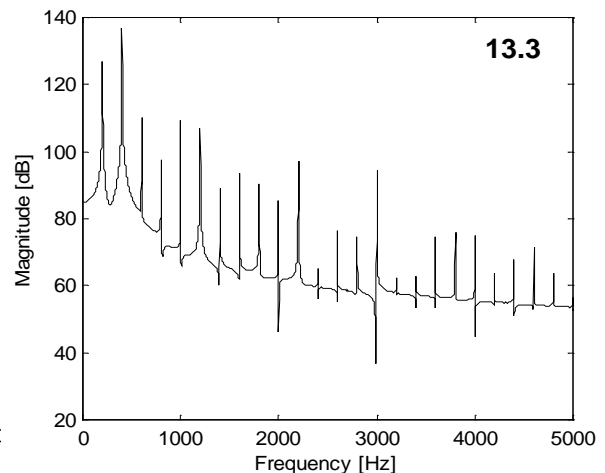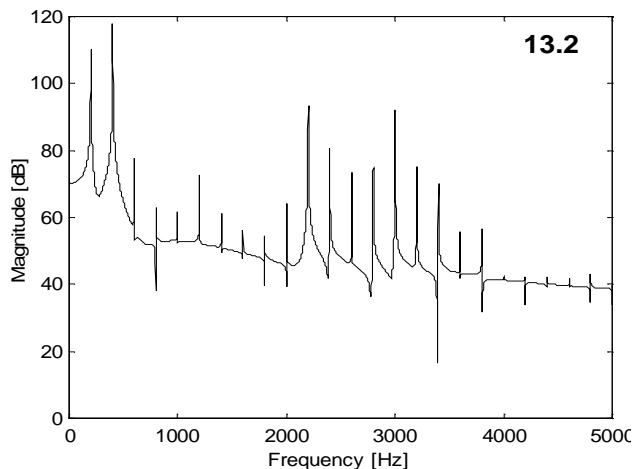

Fig. 13.1-3. Spectra of synthesized /e/. 13.1 shows the original spectrum, 13.2 its spectrum with high pass filter and 13.3 without high pass

III.1.3. Discussion

From physical measurements it became clear, that the Exciter boosts high frequencies and extends bandwidth by introducing linear and nonlinear distortions.

It has been proved that amplifying high frequency regions - typically above 1000 Hz - can result in an increase in speech intelligibility [27]. From a psychoacoustic point of view this finding can be explained with masking. Speech carries little information at frequencies below 1000 Hz, whereas noise (like traffic or indoor car noise) shows its maximum energy at low frequencies. Thus, low frequency noise masks high frequency parts of speech, where much information is carried. Due to the nonlinear behaviour of upward spread of masking, even a small attenuation of low frequencies can cause a distinct 'De-masking' of higher frequencies and thereby result in an improved speech intelligibility.

In the speech-in-noise tests presented in this study, interfering noises always had the same (long term) spectrum as speech. Thus, with low frequency maskers more statistically significant results may be obtained.

The model of sharpness as described in section II.1., predicts an increase in sharpness, if high frequencies are amplified or extended. Although subjects were not asked to judge sharpness directly in the sound quality

experiment, the difference in sound quality often was decribed in terms of sharpness. Looking at the results it seems that additional sharpness is beneficial for musical sounds. Following the theory of perception of nonlinear distortions (section II.2.), the distortions introduced by the Exciter fuse with the original signal to a new percept, which is preferable. But in other cases like speech, the additional sharpness is not suitable and therefore results in a decrease of sound quality. If the original signal already sounds distorted - like 'announcement' - the Exciter's harmonics may fuse with the original distortions, which leads to an even 'more distorted' percept.

Thus, we can state that the Aural Exciter can be regarded as a 'Sharpness Maximizer'. Returning to the definition of a psychoacoustic processor, it is concluded that criterion 2 is fulfilled, since its functional principle takes in account that high frequencies are decisive for sharpness and also criterion 2, since sharpness is enhanced at least in some cases without having detrimental effects on other hearing sensations.
Thus, there's much that is psychoacoustic about the Aural Exciter.

Possible fields of applications outside the music industry are all cases, where signals are transmitted over a band limited channel to a broad band receiver. For example, in hearing aids often a small microphone causes a bandlimited frequency response (about 500 - 5000 Hz), which could be extended to higher frequencies, in particular for middle ear implants, where no speaker is necessary.
Since the Exciter acts - depending in threshold - dynamically, it is possible simulate the 'Lombard-effect' (i.e. high harmonics are emphasized in loud speech) for synthesized speech.

III.2.   Loudness Maximizer

Following the user's manual, with Steinberg's Loudness Maximizer it is possible to increase loudness of normalized sounds without introducing distortions and affecting other hearing sensations like spaciousness or tone color (i.e. sharpness). This is achieved by applying an adaptive algorithm, that controls a combination of slow compression and fast limiting. Depending on the input signal, parameters for compression and limiting are adjusted automatically to obtain an increase in loudness corresponding to a 'desired gain', that is selected by the user. 'Desired gain' is restricted to values smaller than 12 dB. To inform users about the realized increase in loudness, the 'desired gain done' is indicated with LED's. In all psychoacoustic and physical measurements the remaining parameters 'more density' and 'hard/soft' were set to '0'.

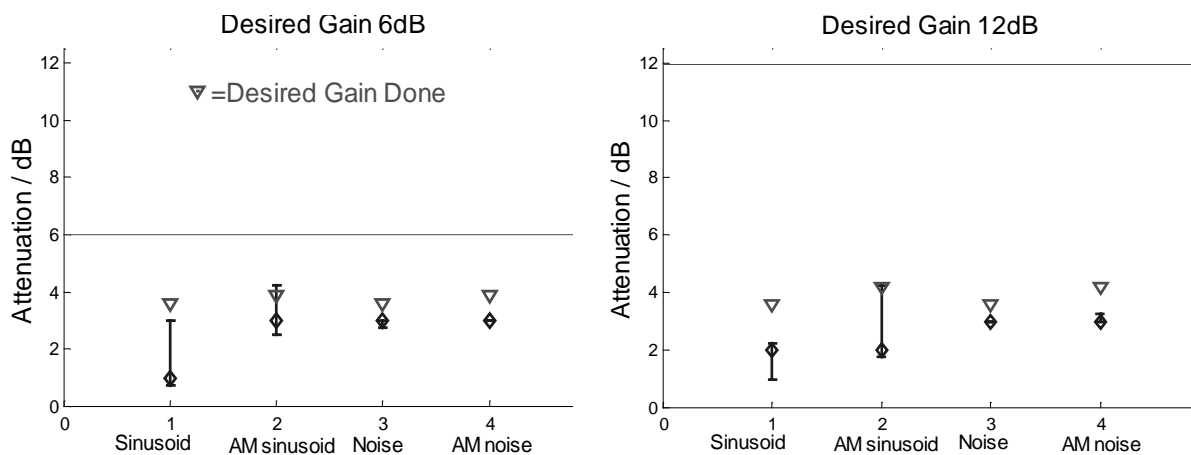### III.2.1. Psychoacoustic Measurements

Fig. 14. Attenuation of maximized (synthetic) sounds for equal loudness

To verify the hypothesis, that the loudness of normalized sounds can be increased by a 'desired gain', level adjustments between original and 'loudness maximized' sounds were carried out, which means that the level of 'maximized' sounds had to be adjusted such that original and 'maximized' sound had the same loudness. Original

sounds were presented diotically (mono) at a (RMS-) level of 70 dB SPL through free-field-equalized DT 48. Subjects could switch as often as they wanted between the original and the maximized stimuli, until they were satisfied with the adjusted level. Before proceeding to the next stimulus subjects were encouraged to mention perceptual differences between the stimuli other than loudness (distortions, dynamics, timbre etc...). All four subjects adjusted the level of each sound four times. Thus, medians and quartiles shown in figures 14 and 15 are calculated from 16 single measurements. The results for various synthetic sounds and a 'desired gain' of 6 dB and 12 dB are depicted in left and right half of figure 14, respectively. Triangles indicate the 'desired gain done'.

The stimuli used in this test were a stationary sinusoid (f = 1 kHz), a stationary white noise, an amplitude modulated sinusoid (f = 1 kHz, modulation frequency = 4 Hz, modulation depth = 30 dB) and an amplitude modulated white noise (modulation frequency = 4 Hz, modulation depth = 30 dB).

While 'desired gain done' is about 4 dB in all cases, psychoacoustic measured gain amounts to a maximum of 3 dB. For a stationary sinusoid only 1 dB is achieved, if 6dB gain are desired. The extremly small interquartile ranges for noises and rather large interquartile ranges for sinusoids are remarkable. This is due to very distinct distortions, which are measurable physically (see next section) and were mentioned by all subjects for all sinusoids. Subjects were confused as to whether they should judge loudness of all spectral components as a whole or just the loudness of the spectral component corresponding to the original sinusoid.

Especially if a desired gain of 12 dB is selected, 'desired gain done' is markedly closer to measured than to desired gains. This indicates that parameters for compression and limiting are chosen by an adaptive algorithm, depending on properties of the input signals.
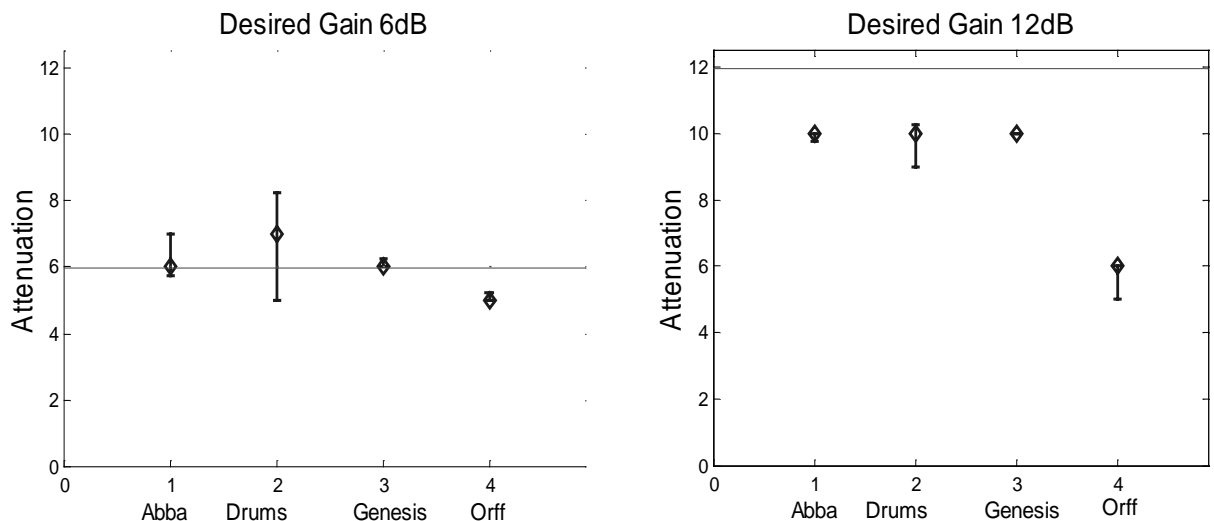

Fig. 15. Attenuation of maximized ('natural') sounds for equal loudness

Only a small increase in loudness, but clearly perceivable distortions are the results with regard to synthetic sounds. Figure 15 shows the results obtained with 'natural' sounds. Besides a song from Genesis and a MIDI-drum loop, two sounds taken from the SQAM CD [28] were used, namely 'Abba' and 'Orff'. A desired gain of 6 dB can be achieved for all stimuli, whereas a gain 12 dB is not realizable. In this case, the necessary attenuation for equal loudness is about 10 dB for 'Abba', 'drums' and 'Genesis', which corresponds to a doubling of loudness. Although the loudness level of 'Orff' is increased only by 6 dB, subjects mentioned - in contrast to the other 'natural' sounds - increased sharpness and annoying distortions, in particular in parts with fast dynamic changes.

In no case was any change in dynamic hearing sensations like fluctuation strength and roughness reported. When critical program material like sinusoids and classical music is maximized, raised sharpness and distortions become audible, whereas pop music seems to be uncritical.

Obviously, the Loudness Maximizer can indeed maximize loudness without affecting other hearing sensations - at least in some special cases.

III.2.2. Physical Measurements

As mentioned above, an adaptive algorithm is the heart of the Loudness Maximizer. Since there is - in contrast to the Aural Exciter - no detailled description of its implementation available, it would be necessary to carry out a huge amount of measurements to reveal its functional principle in general.

Thus, in the framework of this study measurements are restricted to a small number of signals. The block diagram deduced from these measurements therefore is not valid in general, but should be able to demonstrate the essential principles how to maximize loudness.
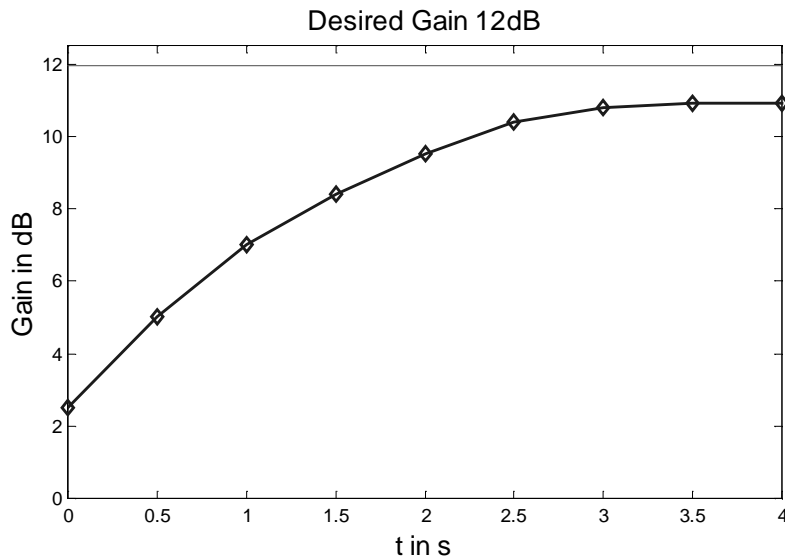
Fig. 16. Gain as a function of time after normalized sinusoid is attenuated abruptly at t=0

Figure 16 shows how gain increases if a normalized sinusoid is attenuated by 12 dB at t=0 s. Gain is calculated from the oszillograms of the maximized and the original sinusoid. According to the hearing experiments in III.2.1. the normalized sinusoid is amplified by about 3 dB. When its amplitude drops, gain is raised very slowly from 3 dB to 11 dB. After 3 seconds gain saturates. This means that the attack time of the compression is about 2 seconds.

Stationary input-output functions were measured with a normalized white noise. The delay between original and maximized sound was calculated from the cross correlation function. Hence, amplitudes - normalized to 1 - of the maximized sound can be plotted against the corresponding original amplitudes. The resulting curve (figure 17) is
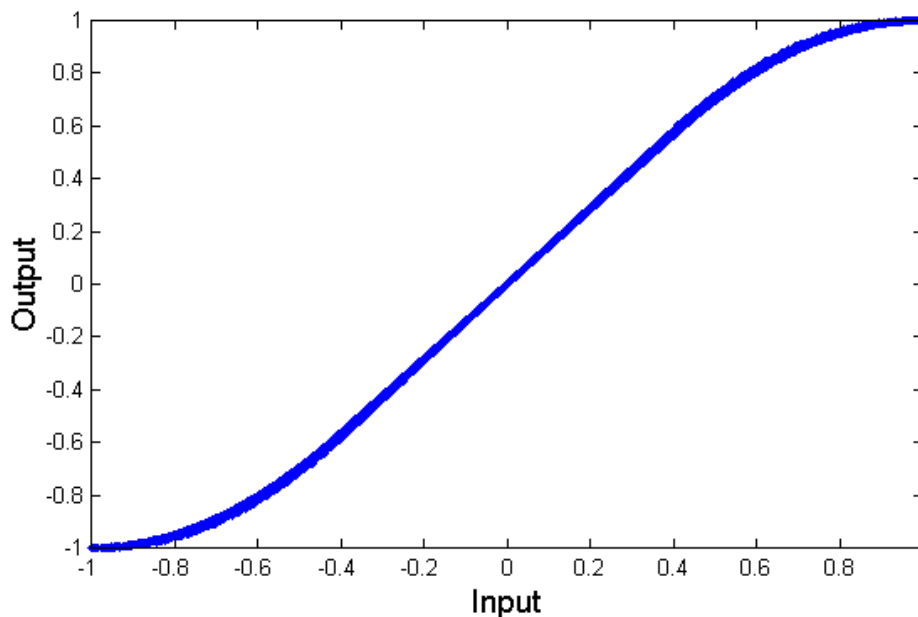
Fig. 17. Input-output function for stationary white noise

linear for a wide range of input amplitudes, in this case corresponding to an amplification of 3 dB. High input values are limited softly to avoid distortions as far as possible (section II.2). The shape of this 'soft knee' can be influenced by 'more density' and 'hard/soft'.
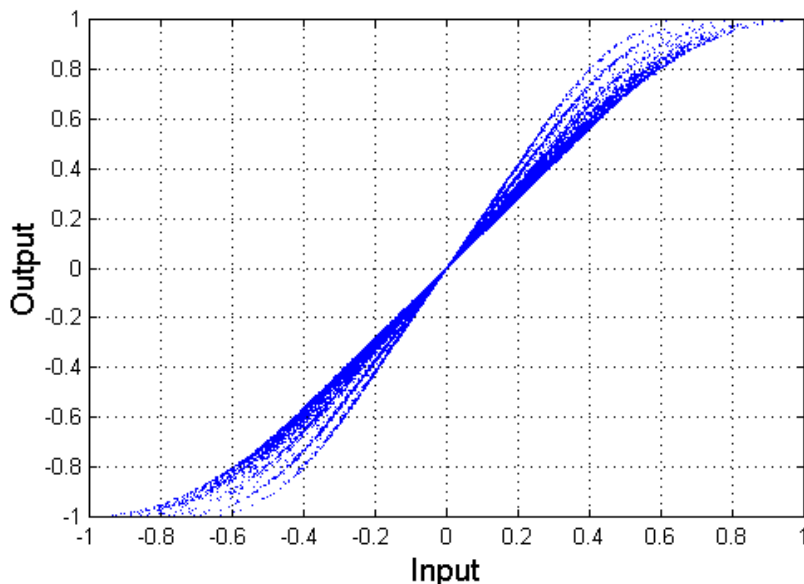

Fig. 18. Input-output function for modulated white noise

Figure 16 demonstrates how this input-output function is adjusted adaptively for a time variant input signal. Although the instantaneous level of the AM white noise varies by 30 dB, the gain applied by the Maximizer only varies between 3 dB and 5 dB. Thus, dynamics are hardly affected. Due to the slow attack time of compression, modulation depth of the maximized sound is nearly the same as of the original sound.

Since the function shown in figure 17 is odd, odd harmonics are introduced (section II.2), when a normalized sinusoid is used as input (figure 19).
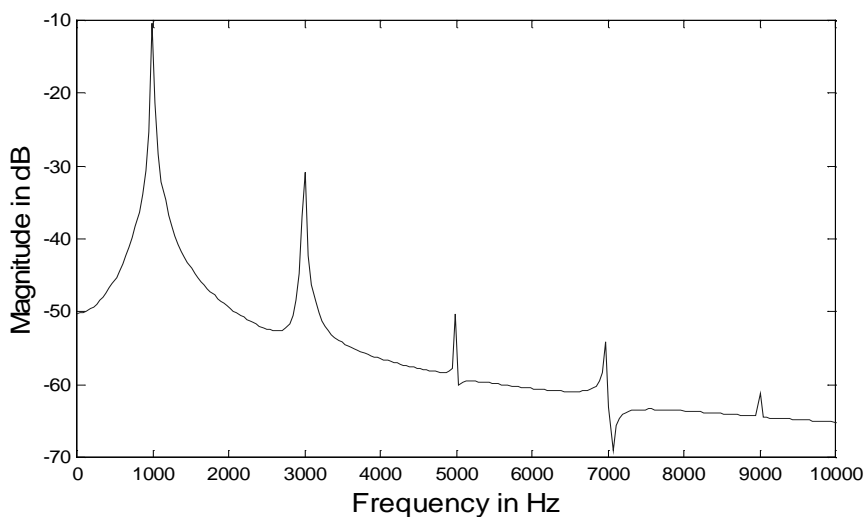

Fig. 19. Spectrum of a maximized 1 kHz tone

The fact that detectable distortions are introduced (see hearing experiment), indicates that the adaptive algorithm does not - or not sufficiently - take into account masking patterns (section II.2).

Based on these measurements we can develop a system, that is able to mimic the Loudness Maximizer at least for the sounds used in this investigation. Since its block diagram is similar to that published by Steinberg and there is no detailed information available, considerations concerning the question what is psychoacoustic about the Loudness Maximizer will based on this block diagram (figure 20).
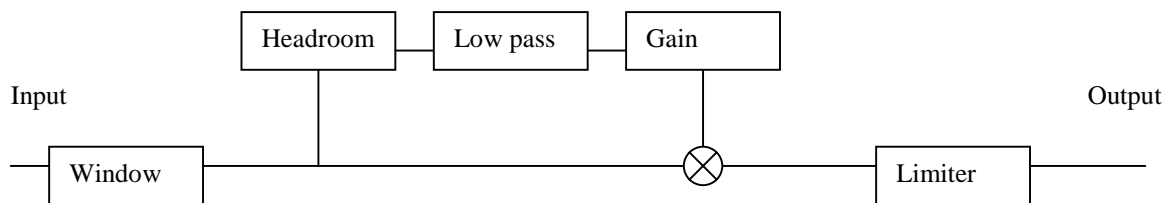
```
                    ┌──────────┐   ┌──────────┐   ┌────────┐
                    │ Headroom │───│ Low pass │───│  Gain  │
                    └────┬─────┘   └──────────┘   └───┬────┘
Input                    │                           │                        Output
       ┌──────────┐      │                           │        ┌──────────┐
───────│  Window  │──────┴───────────────────────⊗───┴────────│  Limiter │──────
       └──────────┘                                           └──────────┘
```

Fig. 20. Block diagram of the 'TUM Maximizer'

Incoming sound is analysed and modified in nonoverlapping rectangular windows (duration: 4 ms). For each window headroom is determined corresponding to the sample with the greatest amplitude. Those headroom values are smoothed by a low pass filter with a cut off frequency of 0.25 Hz. The actual gain, which is applied to the windowed time signal,is then obtained by restricting smoothed headroom to values between 3 dB and 12 dB. If the actual headroom is smaller than actual gain, the time signal is amplified by the amount of actual headroom, which may cause rapid downward changes in gain. This part of the system ('headroom', 'low pass', 'gain') can be regarded as a slow compression.

Calculated gain is also used to determine the input-output function of the following instantaneous limiter. To obtain a smooth curve, the algorithm of Bézier is applied. A Bézier curve is defined by four points: starting point, two control points and endpoint. The starting point is determined by the actual gain, and the end point by the value corresponding to 0 dB FS. By varying the control points similar effects on the curve as by varying 'hard/soft' and 'more density' can be achieved. Concerning the above presented physical measurements, the 'TUM Maximizer' behaves in a very similar though not identical manner to Steinberg's Loudness Maximizer. Informal listening tests suggests that this also true perceptually, at least for the sounds used in the listening experiment.

III.2.3 Discussion

From the results of the hearing experiment (section III.2.1) it can be concluded, that the Loudness Maximizer is capable of raising loudness of normalized sounds without remarkable effects on other hearing sensations like sharpness, fuctuation strength and roughness. This is not true in general, but in special cases, in particular for pop music. Thus, criterion 1 for psychoacoustic processors is fulfilled.

Simply spoken, the Loudness Maximizer acts most of the time as amplifier, resulting in higher critical band levels. Thus, the model of loudness as described in section II.1., can predict the increase in loudness, which will be obtained. The amplification is controlled by an adaptive algorithm, which changes gain very slowly, such that fluctuation strength and roughness are not influenced. This can be achieved by a low pass filter with a cut off frequency of 0.25 Hz, since modulations frequencies lower than 0.5 Hz elicit only very small amounts of fluctuation strength (section II.1.). A soft-knee limiter ensures that annoying distortions are reduced to a minimum (section II.2) for samples that otherwise would exceed 0 dB FS. Due to an analysis window of about half the ear's temporal window (section II.1), abrupt downward changes in gain are smoothed by the human hearing hearing system.

Thus the Loudness Maximizer is a very good example for the application of psychoacoustic knowledge in developing audio signal processing systems.

Summing up we can state that there's much that is psychoacoustic about the Loudness Maximizer.

From a university researcher's point of view, it would be desirable if raised loudness is given in sone, instead of dB. Unfortunately, loudness increases only by a factor of **2**, when level is raised by **10** dB. The latter might be more appealing for marketing reasons, but if so, one should consider to indicate sound pressure or intensity...

Since transmission of audio signals by means of digital channels is very common these days (e.g. mobile phones, digital hearing aids, digital recording & broadcasting), a way to completely exploit available headroom without introducing annoying digital distortion should have many fields of application. In particular, for digital hearing aids, where power consumption is a very critical point, a Loudness Maximizer may be very beneficial, since increasing

loudness by an equivalent level of 10 dB in the digital domain, could save up to 90% of electrical power. Thus, battery lifetime would be prolonged by a factor of 10. The problem is that the Loudness Maximizer increases loudness only in the case of pop music without introducing audible distortions. It is a rather fortunate fact however, that there is an increasing number of young people, who have damaged their ears through listening to loud pop music, but still want to continue to listen to that kind of music...

IV.   SUMMARY

Summing up we can state that there's much that's psychoacoustic about both Aphex's Aural Exciter and Steinberg's Loudness Maximizer. This was shown by carrying out hearing experiments and discussing their results on the basis of psychoacoustic facts and models. While the Loudness Maximizer can be explained by the model of loudness, the Aural Exciter corresponds to sharpness and can be regarded as 'Sharpness Maximizer'. The theory of perception of nonlinear distortions accounts for the astonishing fact that in both cases the use of nonlinearities, which inevitably introduce distortions, frequently does not lead to a deterioration of sound quality.

In this study, psychoacoustic knowledge was used successfully for analysing two audio signal processors, but can - and should - also be applied in developing and describing such systems. In particular, the Loudness Maximizer is a good example for the application of psychoacoustics in the development of new products.

Since both devices are based on rather simple algorithms, but yield markedly psychoacoustic effects, it seems possible that similar algorithms may also be developed in other fields of application, like hearing aids.

ACKNOWLEDGMENTS

REFERENCES

[1] Zwicker, E., Zwicker, U.T., "Audio engineering and psychoacoustics: Matching signals to the final receiver, the human hearing system" , J. Audio Eng. Soc. 39, 115-126 (1991).

[2] Zwicker, E., Fastl, H., "Psychoacoustics". Springer-Verlag 2nd Edition (Berlin Heidelberg New York) (1999).

[3] Terhardt, E., " Calculating virtual pitch", Hearing Research 1, 155-182 (1979).

[4] Widmann, U., Lippold, R., Fastl, H., "A Computer Program Simulating Post-Masking for Applications in Sound Analysis Systems", In: Proceedings of NOISE-CON' 98, Ypsilanti Michi-gan USA, 451-456 (1998).

[5] Stoll, G., Link, M. Theile, G., "Masking -Pattern Adapted Subband Coding: Use of the Dynamic Bit-Rate Margin", J. Audio Eng. Soc. 36, 382, preprint 2585 (1988)

[6] Chalupper, J., "Modellierung der Lautstärkeschwankung für Normal- und Schwerhörige"(Modeling loudness fluctuation for normal and hearing impaired listeners, in German). DAGA2000 (in press).

[7] Terhardt, E., "Fourier transformation of time signals: conceptual revision". Acustica 57, 242-256, 1985.

[8] Mummert, M., "Speech coding by contourizing an aurally adapted spectrogram and its application to data reduction". In German. Ph.D. thesis, Technische Universität München, 1997. For an English description and software download see also http://home.t-online.de/home/Markus.Mummert

[9] Plack, J.P., Moore, B.C.J., "Temporal window shape as a function of frequency and level", J. Acoust. Soc. Am. 87 (5), 2178-2187, 1990.

[10] DIN 45631, "Berechnung des Latstärkepegels und der Lautheit aus dem Geräuschspektrum, Verfahren nach E. Zwicker" (1991)

[11] Rix, A., Hollier, M., "The perceptual analysis measurement system for robust end-to-end speech quality assessment", ". Proc. ICASSP2000, 3, 1515-1518, 2000.

[12] v. Bismarck, G., "Sharpness as an attribute of the timbre of steady sounds", Acustica 30, 159-172 (1974).

[13] Guirao, M., Stevens, S., "Measurement of Auditory Density", J. Acoust. Soc. Am. 36, 1176-1182 (1964)

[14] Fastl, H., " The Psychoacoustics of Sound-Quality Evaluation. In: ACUSTICA - acta acustica 83, 754-764 (1997).

[15] Schmid, W., Chalupper, J., "Pitch Strength: a Psychoacoustical Criterion for Assessing the Quality of Electroacoustical Devices - Why is Phase Response of the Quadratic Difference Tone so important ?" 19. International Convention on Sound Design (Karlsruhe, 1996), VDT, Berlin, 861-874 (1997).

[16] Thornburg, H., "Antialiasing for Nonlinearities: Acoustic Modeling and Synthesis Applications", Proc. ICMC-99 (Beijing) (1999)

[17] Gäßler, G., "Die Grenzen der Hörbarkeit nichtlinearer Verzerrungen bei der Übertragung von Instrumentenklängen", Frequenz 9, 15-25 (1955).

[18] Günthersen, C., "Perception of nonlinear distortion", The Acoustics Laboratory, Technical University of Denmark, report no.34 (1982)

[19] Bregman, A.S., "Auditory Scene Analysis", MIT, Cambridge, Massachusetts (1990)

[20] Schwarz, J., "Pseudo-Psychoakustik", Studio Magazin. Mai 98, 49 (1998)

[21] Herberhold, C., "Gibt es einen Einfluß auf die Sprachverständlichkeit durch Frequenzmodifikation (Exciter) bei Hörgeräteversorgung ?", Klinisch-Praktische Informationen 6, 8. Klinik und Poliklinik für Hals-, Nasen- und Ohrenkranke der Universität Bonn (1998)

[22] Sotscheck, J., "Ein Reimtest für Verständlichkeitsmessungen mit deutscher Sprache als ein verbessertes Verfahren zur Bestimmung der Sprachübertragungsgüte", Der Fernmelde-Ingenieur 4, 5 (1982)

[23] Westra Audiometrie Disc Nr. 2, "Marburger Satzverständlichkeitstest" (1986)

[24] Fastl, H., "A background noise for speech audiometry", Audiol. Acoustics 26, 2-13 (1987).

[25] Aphex, "Aural Exciter Type III Model 250, Operating Guide & Service Manual"

[26] Lindblad, Ann-C., "Influence of nonlinear distortion on speech intelligibility: hearing impaired listeners", Report TA No. 116, Karolinska Institutet, Dept. of Technical Audiology, Stockholm (1987)

[27] Griffiths, J.D., "Optimum Linear Filter for Speech Transmission", J. Acoust. Soc. Am. 43, 81-86 (1964)

[28] EBU, "SQAM - Sound Quality Assessment Material", 1988