INTER-NOISE 2007
28-31 AUGUST 2007
ISTANBUL, TURKEY

Psychoacoustics, sound quality and music

Hugo Fastl[a]
AG Technische Akustik, MMK
Technische Universität München
Arcisstraße 21
D-80333 München
GERMANY

ABSTRACT

In this keynote, relations between psychoacoustics, sound quality, and music are discussed.
In the domain of psychoacoustics, a Dynamic Loudness Model (DLM) is presented which can simulate loudness perception of sounds with strong temporal structure for both normal hearing persons and persons with (slight) hearing deficits. Recent results on the hearing sensation pitch strength will be presented and discussed in view of applications both in sound quality design and music.
Sound quality is assessed on the one hand for products where it is a genuine feature like musical instruments, and on the other hand where it is just a by-product like Diesel engines or closing car doors.
It is shown that the psychoacoustic magnitudes loudness and sharpness can account for largely different aspects like the annoyance of snoring sounds on the one hand and the sound quality of grand pianos on the other hand.
The concept of psychoacoustic annoyance based on noise evaluation is contrasted with the concept of sensory pleasantness, anchored in musical acoustics.
For questions of sound quality design, an approach incorporating elements of decision tree studies is proposed.
The strong ties between psychoacoustics, sound quality and music have been advocated already about 150 years ago by Helmholtz. He postulated that musical consonance is governed by the absence of roughness. These days, roughness is one of the major ingredients for sound quality design. Also, musical dynamics from ppp to fff is reflected in modern category scaling procedures used in psychoacoustics as well as sound quality rating. Moreover, A-weighted level can be misleading with respect to loudness perception for both musical and technical sounds. Therefore, a loudness-thermometer is contrasted with a level-thermometer, including possible cognitive effects of recognizing the sound source.

1 INTRODUCTION

The goal of the scientific field of psychoacoustics is to establish firm relations between sounds which are well defined in the physical domain and the hearing sensations elicited by these sounds. Probably the oldest experiments in psychoacoustics have been performed in a musical context: For example around 500 B.C., Pythagoras studied musical consonance and dissonance. The experimental setup was a monochord, i. e. a string stretched along a board with a bridge, the position of which could be varied. The physical input to these very early psychoacoustic experiments is the ratio of the length of the parts of the string which are divided by the bridge. Pythagoras found out that if this ratio represents small integer numbers

[a] Email address: fastl@mmk.ei.tum.de

like 2 : 1 (octave), 3 : 2 (fifth), 4 : 3 (fourth) etc., consonant musical intervals are obtained. Although with modern digital technology the generation and presentation of sounds has been considerable refined, the aim of psychoacoustic experiments is still the same: To trace as much as possible the physical basis of auditory perception.

This concept has proven very successful over the years. For example, early psychoacoustic studies performed at Bell Labs (cf. Allen 1996) helped to significantly improve the communication via telephone. In this context it should be mentioned that the scientific field psychoacoustics profited substantially from the research of telecommunication engineers like the Nobel-Laureate Georg von Békésy (1899-1972) or the *Doyen* of psychoacoustics in Germany, Eberhard Zwicker (1924-1990). Since engineers typically do not stop their research in basic science, but usually are looking for practical applications of the results achieved, it is not astonishing that data from basic psychoacoustics more and more found their way into practical applications (cf. Fastl and Zwicker 2007).

Of course the ties of applied psychoacoustics were and still are very strong to musical acoustics. In addition, these days many applications of psychoacoustics are found in audiology (e.g. Kollmeier 1996). One of the largest fields of practical applications for knowledge from both psychoacoustics and music is sound quality evaluation and sound quality design (e.g. Fastl 1997). For example for household appliances, the decision to buy a specific product depends on the following factors:

1. efficiency, 2. price, 3. sound, ... 7. design, ...

Hence, the sound produced by a product can play an important part, not only for household appliances (e.g. Bayraktar and Belek 2006), but also for other products. Despite the fact that the (optical) design of cars is much more important than that of household appliances, the automotive industry affords large investments to improve the sound quality of the interior sound of cars (e.g. Hashimoto 2000). More recently, also the exterior sound of cars gets more and more relevance. In addition to these well known applications of psychoacoustics also more exotic examples exist like the sound quality when opening a beer bottle or when munching crackers.

In this paper, first some recent developments in psychoacoustics are discussed, second examples for sound quality rating as well as sound quality engineering are given, and finally it is illustrated that psychoacoustics and sound quality evaluation still show strong ties to music.

2 PSYCHOACOUSTICS

From more recent developments in psychoacoustics relevant for sound quality evaluation and music perception, two examples shall be described: On the one hand, a Dynamic Loudness Model (DLM), and on the other hand, the hearing sensation pitch strength.

2.1 Dynamic Loudness Model (DLM)

Current loudness models as described for example in DIN 45631 or ISO 532 assess the loudness of stationary sounds as perceived by normal hearing persons. Since, however, most sounds of everyday life are non-stationary, a Dynamic Loudness Model (DLM) was realized (Chalupper and Fastl 2002). The main advantage of the Dynamic Loudness Model is that it can assess loudness perception of normal hearing persons as well as persons with (slight) hearing deficits for both stationary and non-stationary sounds.

Figure 1 shows a block diagram of the Dynamic Loudness Model.

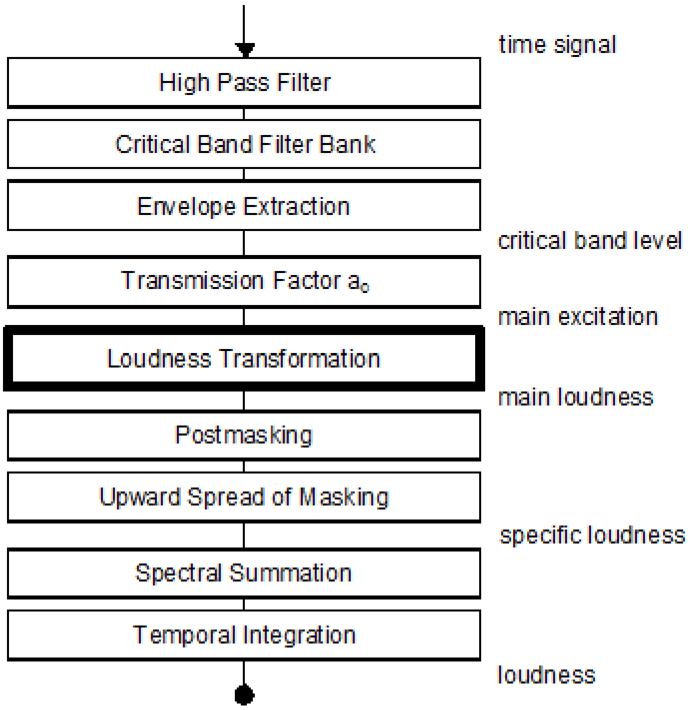

Figure 1: Block diagram of the Dynamic Loudness Model (DLM) proposed by Chalupper and Fastl 2002.

The blocks displayed in Figure 1 illustrate that the DLM is a typical Zwicker-type loudness model with critical band filters, post-masking, upward spread of masking as well as spectral summation and temporal integration. An important feature of the Dynamic Loudness Model is represented by the block "Loudness Transformation": When describing loudness perception of persons with hearing deficits in comparison to loudness perception of persons with normal hearing, only this block has to be changed. In former loudness models for persons with hearing deficits, it was assumed that broader critical band filters, shallower post-masking curves as well as reduced temporal integration would be necessary. However, it turned out that when switching in signal processing from the level domain to the loudness domain, only the modified loudness transformation of persons with hearing deficits has to be taken into account.

The difference in loudness processing between normal hearing persons and a person with a hearing deficit is illustrated in Figure 2.

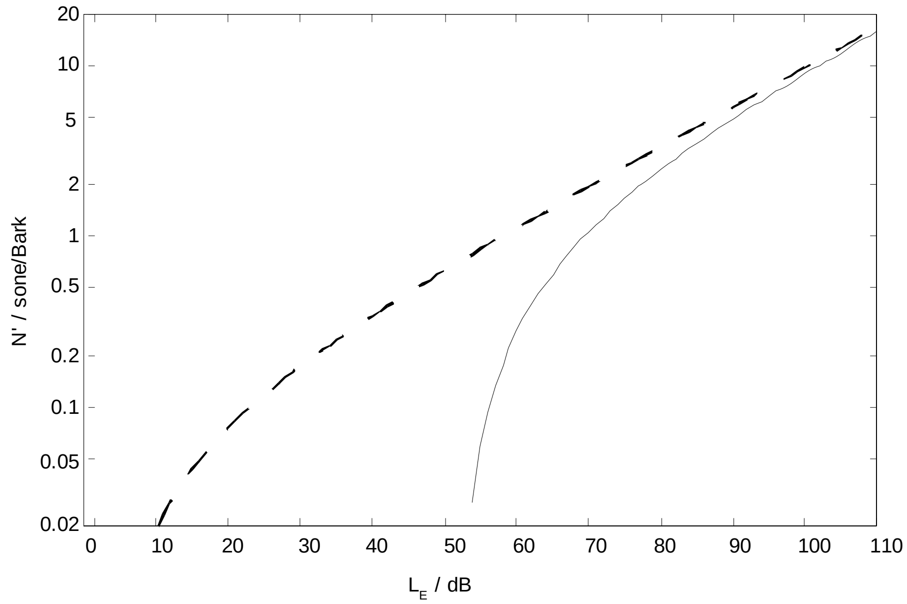

Figure 2: Loudness transformation for normal hearing persons (dashed) and a person with about 40 dB hearing loss (solid) after Chalupper and Fastl 2002.

The dashed line in Figure 2 indicates that over a large range of levels the rule of thumb "10 dB more means double loudness" is valid. However, for a person with hearing deficit, an increase in level by 10 dB can produce an increase in specific loudness of about a factor of 20. This phenomenon, which is called *recruitment*, is well known from everyday life: When talking to a person with hearing deficit, frequently the person will ask you to "speak up". If you then increase your voice only slightly, the hard of hearing person usually tells you "don't shout at me". This behavior is typical for persons who suffer from recruitment.

As an example, Figure 3 shows loudness scaling of narrow-band noises by normal hearing persons in comparison to a person with hearing deficits, including recruitment at high frequencies.

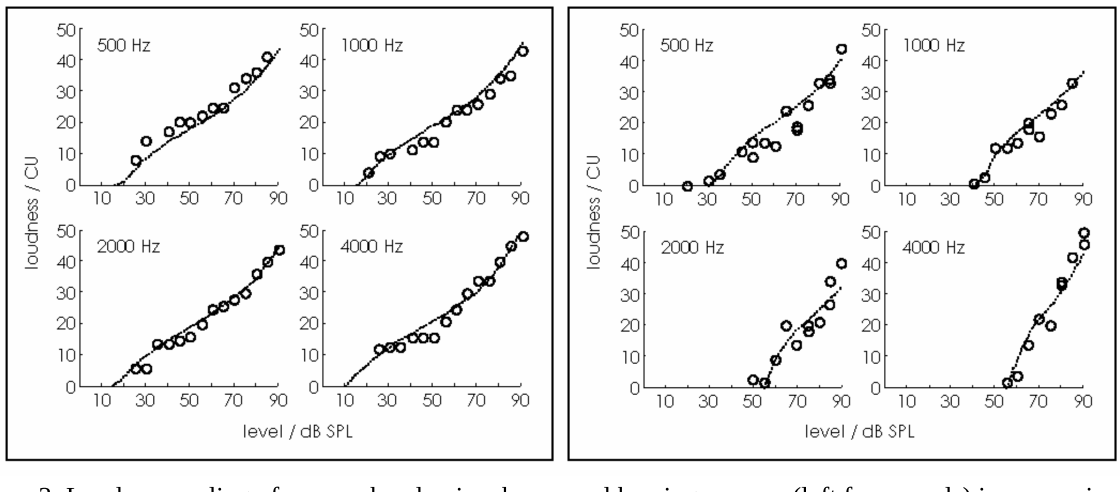

Figure 3: Loudness scaling of narrow-band noises by normal hearing persons (left four panels) in comparison to a person with hearing deficits at high frequencies (right four panels). Circles: subjective loudness evaluations, Curves: predictions by the Dynamic Loudness Model. (after Chalupper 2002)

The data displayed in the left four panels of Figure 3 illustrate loudness evaluation of narrow-band noises at different center frequencies by normal hearing persons using a method of category scaling (cf. Hellbrück and Ellermeier 2004). The right four panels indicate loudness evaluations of a person with hearing deficits at high frequencies. At 500 Hz, this person shows almost normal loudness scaling. However, at high frequencies of 2000 Hz and in particular at 4000 Hz, loudness evaluations typical for recruitment show up: A narrow-band noise at 4000 Hz with about 60 dB is perceived as very faint (5 CU), but at 90 dB as very loud (50 CU). The left panel for 4000 Hz shows that normal hearing persons rate the narrow-band noise with 90 dB also as very loud with 50 CU. This means that the loudness sensation of the person with hearing deficits “catches up” at higher levels.

For sound quality design, it is very important, to take into account loudness perception of persons with (slight) hearing deficits. On the one hand, elder persons frequently show reduced hearing ability at higher frequencies. On the other hand, an increasing number of younger persons unfortunately also shows hearing deficits at higher frequencies, frequently due to extremely loud leisure activities.

2.2 Pitch Strength

As already mentioned, experiments on pitch (height) belong to the oldest topics assessed in psychoacoustics. However, at same pitch height, sounds can produce pitches of rather different strength. For example a pure tone elicits a strong, definite pitch, whereas a low pass noise produces only a faint pitch. This magnitude is assessed by the hearing sensation pitch strength (cf. Fastl and Stoll 1979).

Figure 4 shows different sounds which produce approximately the same pitch but differ strongly in pitch strength.

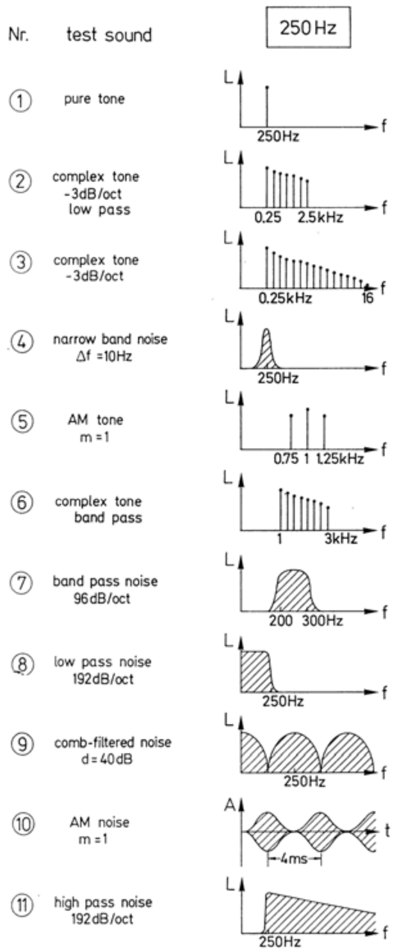

Figure 4: Excerpt from the stimuli used for the experiments on pitch strength (Fastl and Stoll 1979).

The sounds used include line spectra as well as stochastic signals. The perceived pitch height corresponds for all signals approximately to the pitch produced by a pure tone at 250 Hz. The relative pitch strength of the signals illustrated in Figure 4 is displayed in Figure 5.

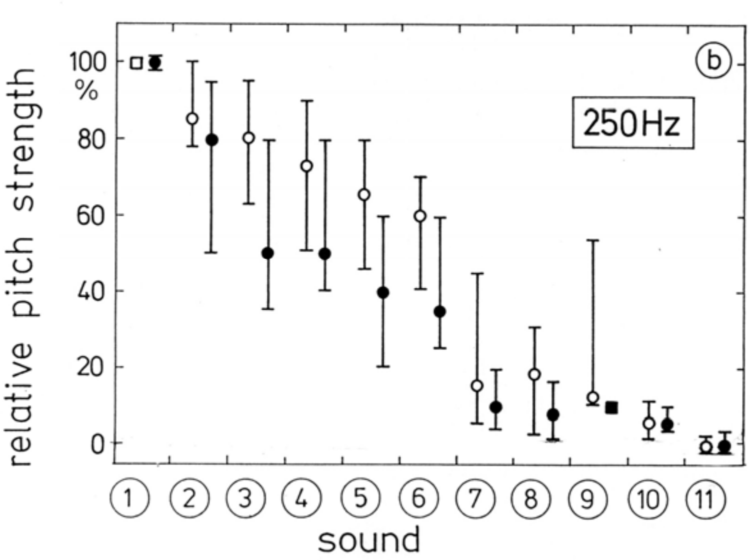

Figure 5: Relative pitch strength of the signals illustrated in Figure 4.

The data plotted in Figure 5 indicate that the largest pitch strength is produced by a pure tone. In comparison to sounds with line spectra, stochastic signals (7 to 11) produce only very faint pitches of about 10 % or less in comparison to the pitch strength of a pure tone. However, one exception is sound (4), a narrow-band noise with extremely small bandwidth (10 Hz) which elicits a pitch strength of about 60 % compared to the pitch strength of a pure tone. Since a narrow-band noise at such small bandwidth sounds like a pure tone which is modulated in amplitude and frequency, this behavior is not unexpected.

Figure 6 illustrates the block diagram of a model of pitch strength developed by Fruhmann (2006a)

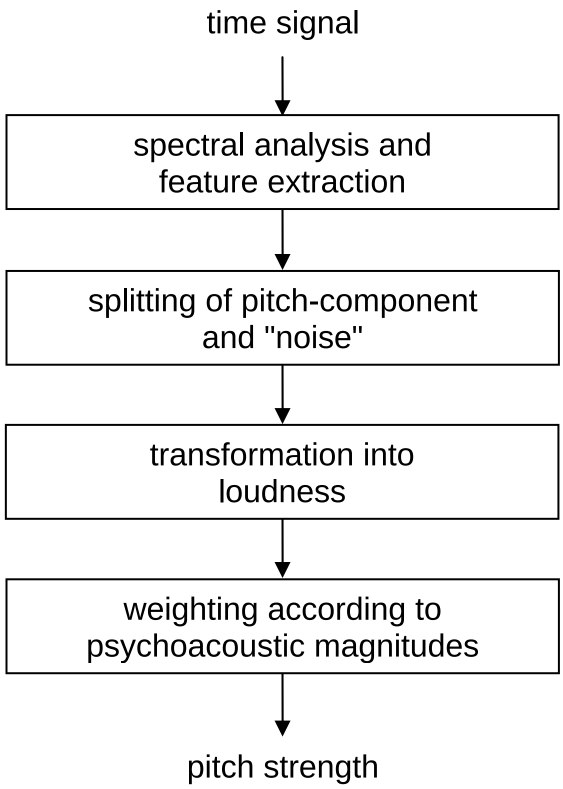

Figure 6: Block diagram of a model of pitch strength (after Fruhmann 2006a).

Fruhmann's model of pitch strength starts with a FTT-analysis (Terhardt 1985) and splits the pitch of the fundamental from the remaining "noise". After a transformation into loudness and weighting by the psychoacoustics magnitudes fluctuation strength and roughness, the pitch strength is obtained.

The signal processing in the pitch strength model of Fruhmann is illustrated in more detail in the following three figures. Figure 7 shows for a pure tone at 1 kHz (8.5 Bark) the FTT-spectrum, the loudness pattern, and the loudness-time function.

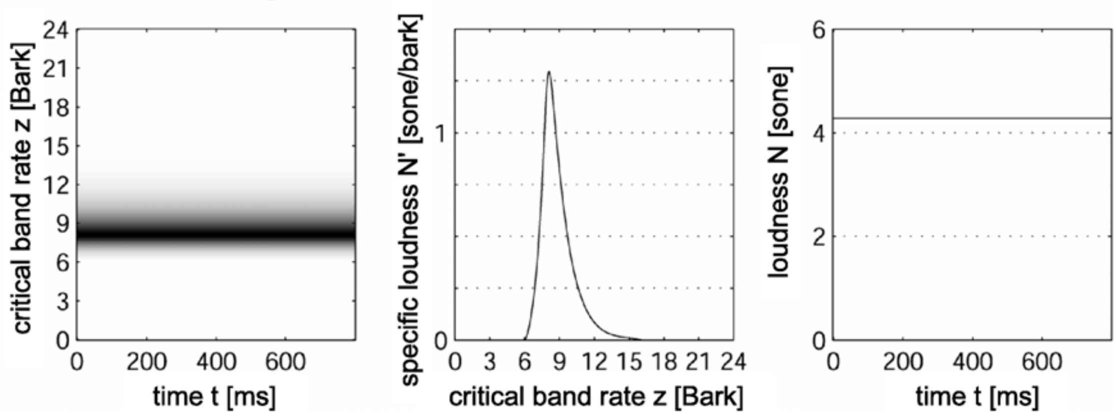

Figure 7: Illustration of signal processing for a pure tone at 1 kHz (8.5 Bark) in the model of pitch strength. FTT-spectrum (left), loudness pattern (center), and loudness-time function (right) (from Fruhmann 2006b).

The FTT-spectrum displayed in Figure 7 shows a clear maximum at 8.5 Bark, the critical band rate corresponding to the 1 kHz pure tone. The loudness pattern is very selective, and the loudness-time function is constant at about 4 sone, indicating that the level of the pure tone is around 60 dB.

Figure 8 illustrates the processing of a harmonic complex tone with 500 Hz fundamental frequency in the model of pitch strength.

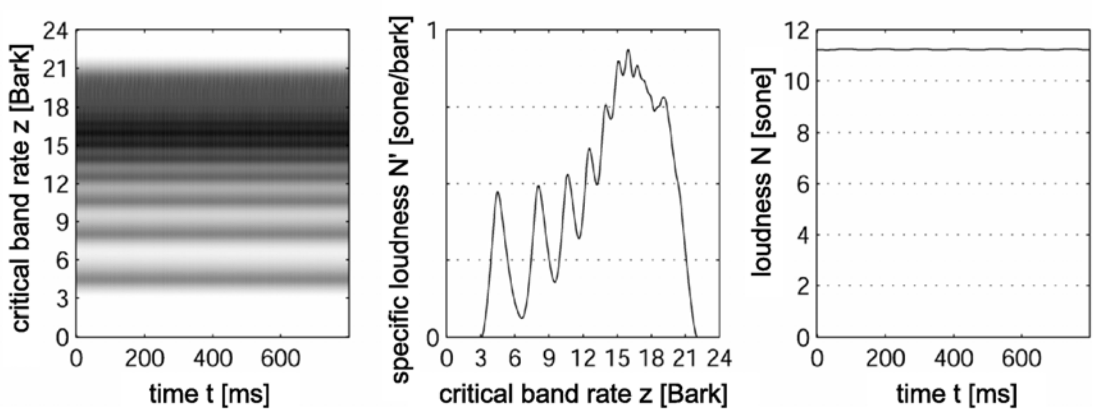

Figure 8: Illustration of signal processing for a harmonic complex tone with 500 Hz fundamental frequency in the model of pitch strength. FTT-spectrum (left), loudness pattern (center), and loudness-time function (right) (from Fruhmann 2006b).

The FTT-spectrum (left) clearly shows the harmonic structure of the harmonic complex tone and the fact that about the lowest six harmonics are resolved by the hearing system. This effect is also seen in the loudness pattern (center): The lower harmonics are resolved, and the higher, unresolved harmonics represent in the model of pitch strength “noise”. The loudness-time function (right) also shows no temporal variation, but a larger value (about 11.5 sone) than in Figure 7, since total loudness is proportional to the area in the loudness pattern (Zwicker 1960).

Figure 9 illustrates the signal processing in the model of pitch strength for a narrow-and noise with 20 Hz bandwidth (upper panels) or a narrow-band noise with 200 Hz bandwidth (lower panels). Both noises are centered at 1 kHz.

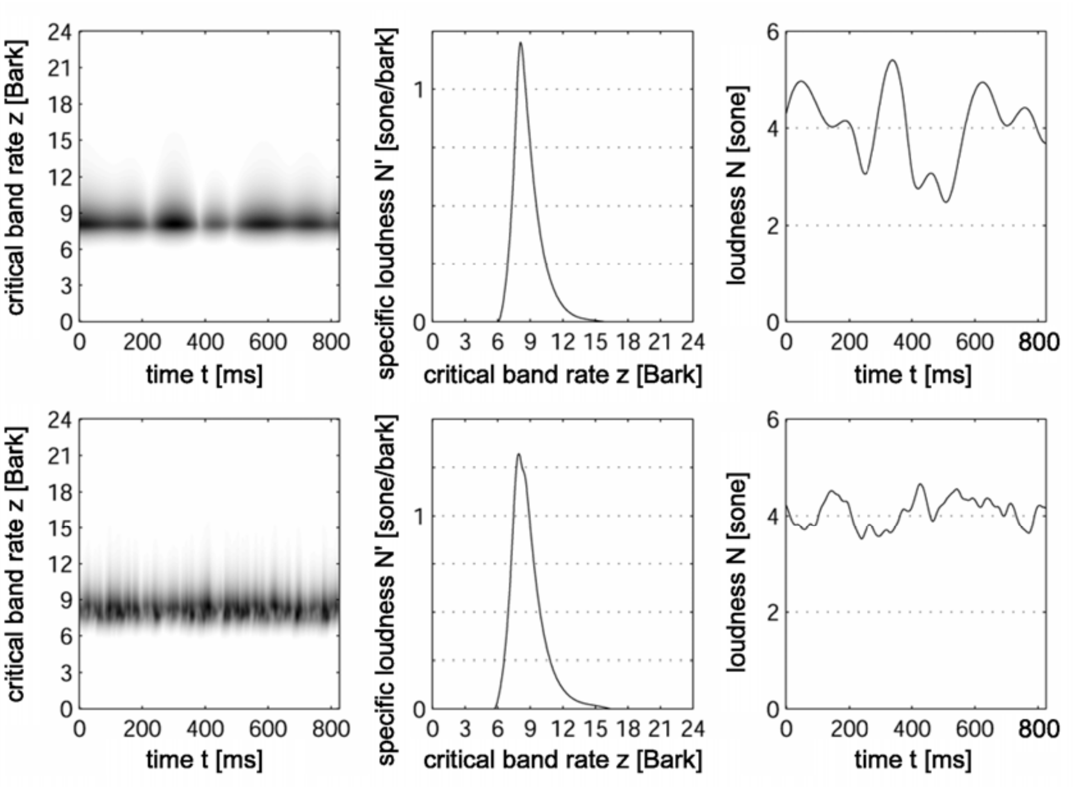

Figure 9: Illustration of signal processing for narrow-band noises centered at 1 kHz with 20 Hz bandwidth (upper panels) or 200 Hz bandwidth (lower panels). FTT-spectra (left), loudness patterns (center), and loudness-time functions (right) (from Fruhmann 2006b).

The FTT-spectra in the left part of Figure 9 illustrate the stochastic nature of the narrow-band noises considered. The loudness patterns in the center reflect the larger bandwidth of the 200 Hz wide noise. The right panels illustrate that the fluctuation of the loudness- time function is larger and slower for the narrow-band noise with 20 Hz bandwidth (upper right panel) than for the narrow-band noise with 200 Hz bandwidth (lower right panel).

The data displayed in Figure 10 enable a comparison of subjective evaluations (Fastl 1989) and predictions by the model of pitch strength (Fruhmann 2006a) for the dependence of pitch strength of pure tones on relevant stimulus parameters.

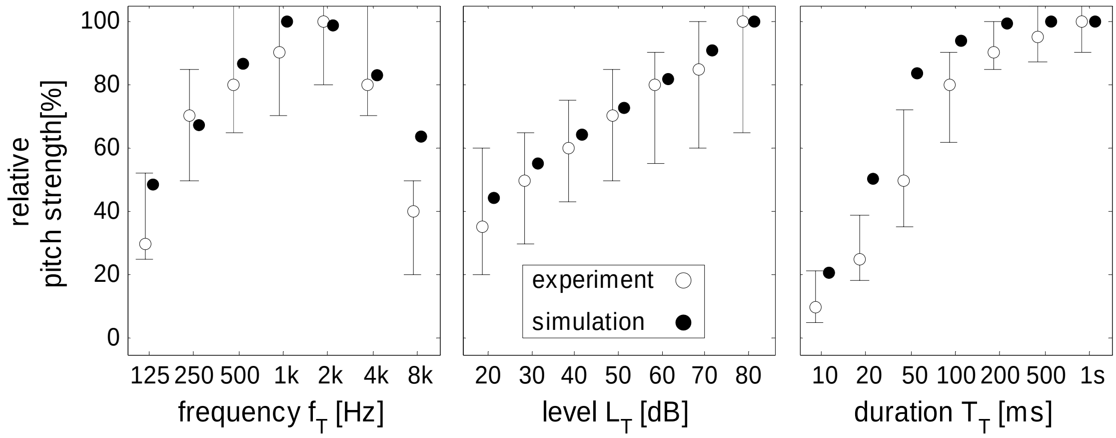

Figure 10: Dependence of relative pitch strength of pure tones on frequency (left), level (center), and duration (right). Subjective data from Fastl 1989 (unfilled circles), and predictions by the pitch strength model of Fruhmann 2006a (filled circles)

The data plotted in Figure 10 indicate that the model of pitch strength nicely accounts for the dependence of the pitch strength of pure tones on relevant stimulus parameters. In many cases, the predicted values of relative pitch strength (filled circles) are near the medians (unfilled circles) or at least within the inter-quartile ranges of the subjective data.

The data displayed in Figure 11 allow a comparison of subjective evaluations and predictions of pitch strength of narrow noise bands as a function of bandwidth.

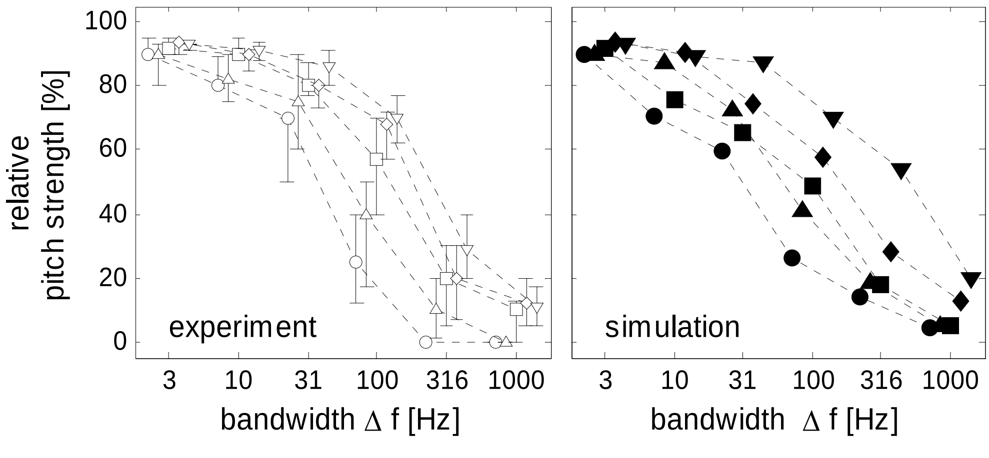

Figure 11: Pitch strength of narrow-band noises as a function of bandwidth. Subjective evaluations (Wiesmann and Fastl 1991, left panel) and predictions by the model of pitch strength (Fruhmann 2006a, right panel)

The data plotted in the left panel of Figure 11 indicate that the pitch strength of narrow-band noises decreases with increasing bandwidth. However, at one and the same bandwidth (e.g. 100 Hz) pitch strength of narrow-band noises also depends crucially on their center frequency: While for a center frequency of 250 Hz (circles), relative pitch strength reaches only about 25 %, at 4000 Hz center frequency (downward pointing triangles), relative pitch strength amounts to about 70 %. Both the decrease of pitch strength with increasing bandwidth and the dependence of pitch strength of a narrow-band noise with constant bandwidth on the center frequency is predicted by the model of pitch strength (right panel).

Figure 12 illustrates a practical application of pitch strength in noise evaluation. For the indoor noise of high speed trains, the effects of the motor or of corrugated rails have been studied (Patsouras et al. 2002). A typical result of corrugated rails is that the third octave band around 1250 Hz is elevated in comparison to neighboring one third octave bands. Effects of the motor usually produce elevations of the third octave band around 630 Hz (e.g. Patsouras 2003). The subjective evaluation is illustrated in Figure 12 by unfilled symbols with inter-quartiles, the predictions by the model of pitch strength are indicated by filled symbols.

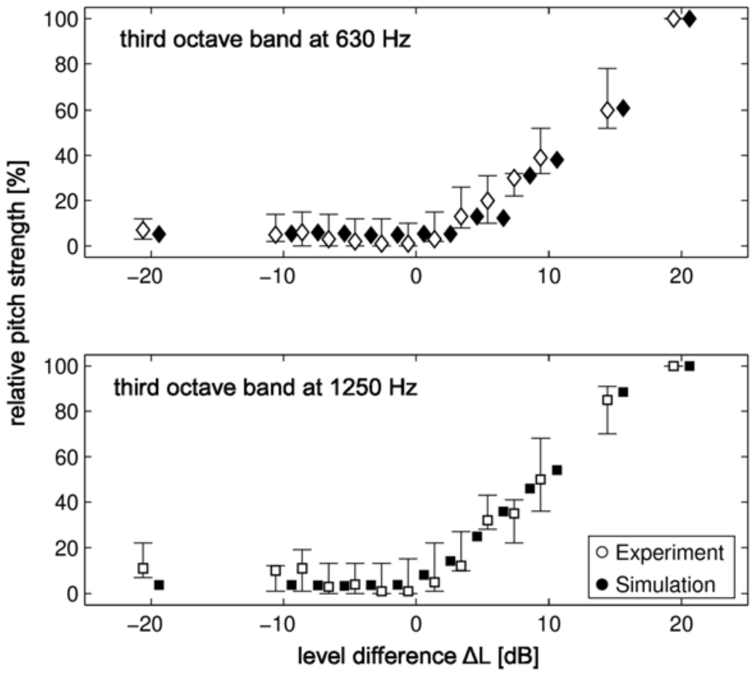

Figure 12: Comparison of subjective evaluation (unfilled symbols) and prediction by the model of pitch strength (filled symbols) of noises due to effects of the motor (upper panel) or corrugated rails (lower panel). Subjective data from Patsouras et al. (2002), simulation from Fruhmann (2006b)

The comparison of unfilled and filled symbols in Figure 12 reveals that the model of pitch strength nicely accounts for the effects of the motor or corrugated rails on sound quality. In the experiments, the largest values of relative pitch strength were found for a third octave band which is 20 dB above the neighboring third octave bands. About half the pitch strength is obtained for a third octave band which is about 10 dB above neighboring bands and almost no pitch strength if the third octave band is at same height as its neighbors.

Figure 13 illustrates an example of the application of pitch strength in musical acoustics. It is well known in musical circles that professional players of timpani prefer natural heads from calf skin to synthetic heads because of their “pure sound”. In psychoacoustic experiments (Fastl and Fleischer 1992) it could be shown that the pitch strength of sounds from timpani with natural heads tends to be larger than the pitch strength of sounds from the same timpani with synthetic heads.

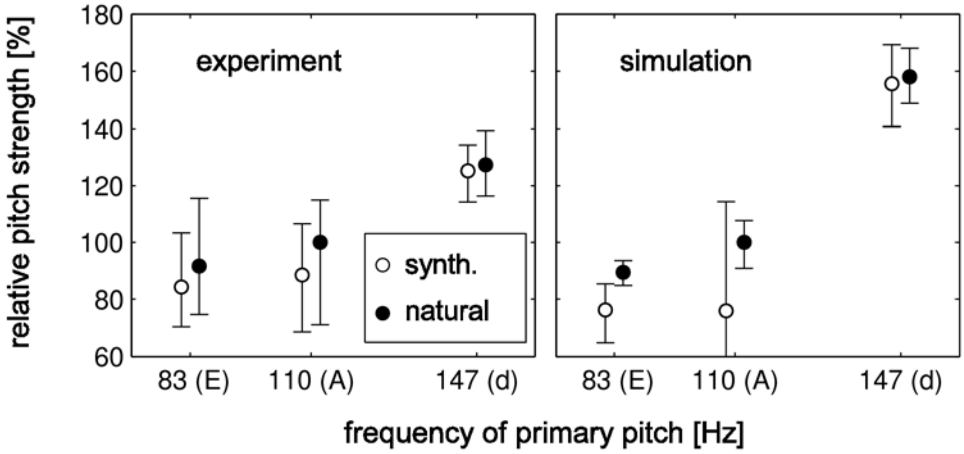

Figure 13: Relative pitch strength of sounds from timpani with natural head (filled symbols) in comparison to sounds from the same timpani with synthetic head (unfilled symbols). Subjective evaluations after Fastl and Fleischer 1992 (left), and predictions by the model of pitch strength from Fruhmann 2006b (right).

The data plotted in Figure 13 indicate the tendency that the pitch strength of timpani with natural head may be slightly larger than the pitch strength of the same timpani with synthetic head. The right panel in Figure 13 shows that the calculations from the pitch strength model point in the same direction. However, it has to be realized that the sounds were evaluated by subjects with experience in psychoacoustics, but not professional timpani players. Therefore, the variations in the generation of the sounds as well as the subjective evaluations were rather large compared to the data expected for professional musicians.

3 SOUND QUALITY

Sound quality ratings can be divided into two groups: The first group deals with the rating of sound quality of products, for which sound quality is a genuine feature like musical instruments or loudspeakers. The second group encompasses rating of sound quality of products where it is just a by-product like sound quality of diesel engines or closing car doors.

3.1 Sound quality as a genuine feature

As an example, where sound quality is a genuine feature of the product, electric guitars are considered. Since the strings of electric guitars perform bending waves, the resulting spectra are not completely harmonic (cf. Zollner 2007). Figure 14 shows typical values for the index of inharmonicity for the strings of an electric guitar.

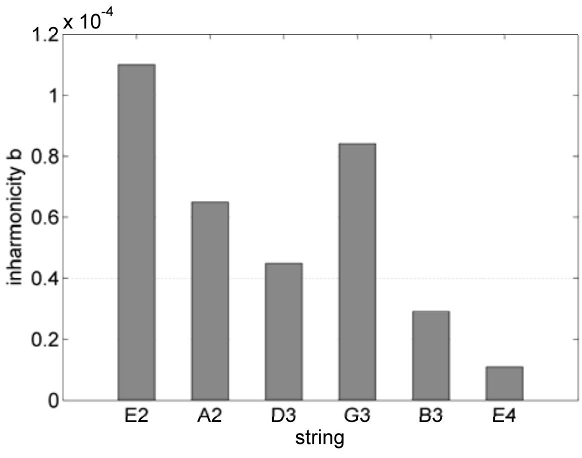

Figure 14: Typical values for the inharmonicity of the strings of an electric guitar (after Zollner, 2007)

The data displayed in Figure 14 clearly reveal that the largest inharmonicity occurs at the lowest string of the electric guitar tuned to E_2. For higher strings (A_2, D_3) the inharmonicity decreases, but increases again for the string G_3. This behavior is due to the fact that the lower three strings of an electric guitar are wound strings, and the diameter of the string kernel decreases. However, for the G_3 string, which is an unwound string, the diameter is again larger, leading to larger inharmonicities.

Sounds of an electric guitar were synthesized and evaluated with respect to sound quality (Völk et al. 2006). Figure 15 gives an example for the preference of short melodies when the string sounds are realized by perfectly harmonic sounds versus slightly inharmonic sounds.

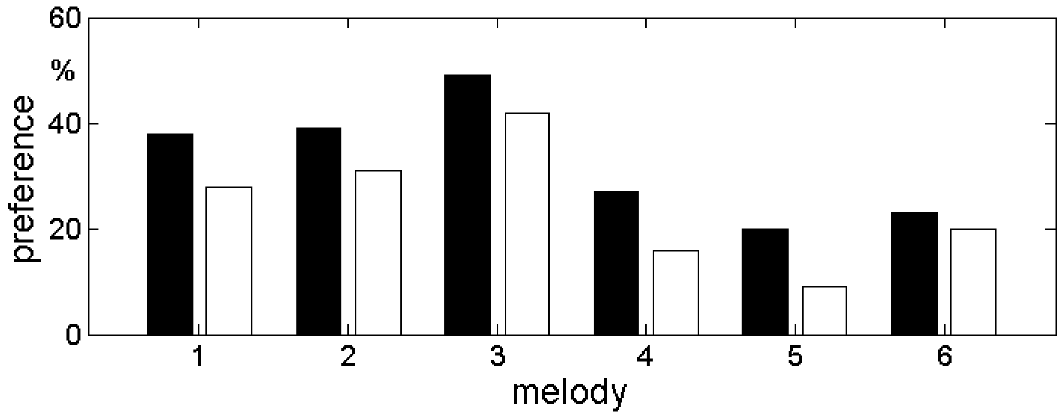

Figure 15: Comparison in sound quality between short melodies played on a simulated electric guitar using strictly harmonic (unfilled columns) versus slightly inharmonic (filled columns) string sounds (Völk et al. 2006).

The data displayed in Figure 15 reveal that slightly inharmonic guitar sounds are somewhat preferred compared to strictly harmonic guitar sounds. This result means that the inharmonicity of the strings of electric guitars is *not* considered by musicians as an inevitable physical flaw which should be reduced by “strictly harmonic” sounds. On the contrary, slightly inharmonic string sounds are typical for an electric guitar, and usually produce a better sound quality in a musical context.

3.2 Sound quality as a by-product

A very typical example of sound quality rating, where sound quality is a by-product, is the sound of closing car doors. In the show room of car dealers, the sound of closing doors is of decisive importance, since a good quality of the sound produced by closing a car door subconsciously is related to a good quality of the whole vehicle. Because these effects are of great relevance for an international market, subjective ratings of sounds from closing car doors were performed both with Japanese and German subjects. Excerpts from a larger study by Sonoko Kuwano (Kuwano et al. 2004) are illustrated in the following three figures.

Figure 16 shows the semantic differential for the sound of a closing car door, where the subjects felt it would stem from a luxurious sedan.

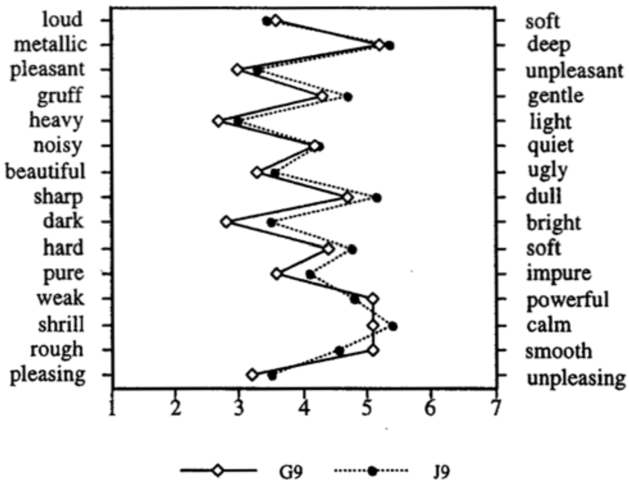

Figure 16: Semantic differential for the sound of closing a car door where the subjects felt it would stem from a luxurious sedan. Filled symbols: Japanese subjects; unfilled symbols: German subjects (from Kuwano et al. 2006)

The data displayed in Figure 16 show that the door sound of a luxurious sedan is characterized by adjectives like deep, pleasant, gentle, heavy, dark, and so forth. The data of Japanese and German subjects show close correlation.

Results displayed in Figure 17 illustrate the rating of the door sound quality for a car which subjects rated as an economy sedan.

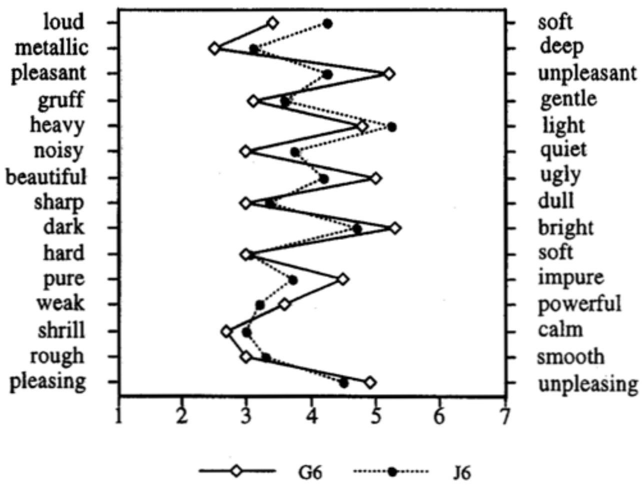

Figure 17: Semantic differential for the sound of closing a car door where the subjects felt it would stem from an economy sedan. Filled symbols: Japanese subjects; unfilled symbols: German subjects (from Kuwano et al. 2006)

The door sounds of the economy sedan can be characterized by adjectives like metallic, unpleasant, gruff, light, noisy, ugly, sharp, and so forth. Again, ratings by Japanese and German subjects are in agreement.

Figure 18 shows the semantic differential for the sounds of a closing door from a pick up truck.

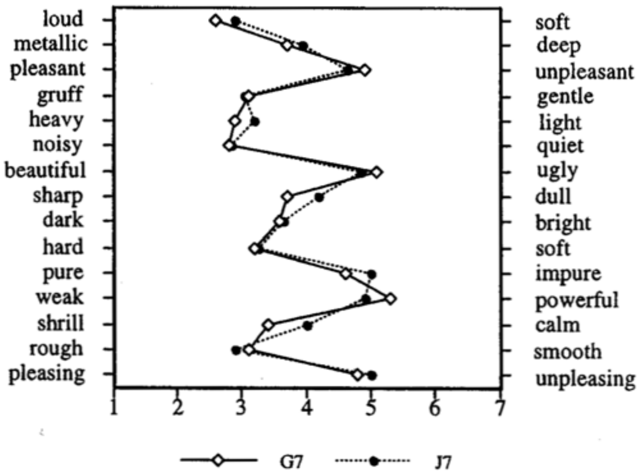

Figure 18: Semantic differential for the sound of closing a car door where the subjects felt it would stem from a pick up truck. Filled symbols: Japanese subjects; unfilled symbols: German subjects (from Kuwano et al. 2006)

The typical sound for the closing door of a pick up truck is loud, unpleasant, gruff, heavy, noisy, ugly, and so forth. Again data for Japanese subjects and German subjects show good agreement.

When comparing the profiles displayed in Figure 16 to Figure 18 it becomes clear that the rating for a luxurious sedan (Figure 16) is more or less inverse to the rating for an economy sedan (Figure 17). For a pick up truck (Figure 18), the evaluation of the door sound follows a rather different pattern. Obviously, the expectation of subjects for the door sounds of sedans is very different from the expectation of the door sound from a pick up truck.

Another example of sound quality rating, where sound quality is just a by-product with considerable practical relevance is the sound quality evaluation of a Diesel motor (e.g. Hastings and Davies 2006). In modern Diesel motors, a conflict between fuel consumption and noise is evident: For “hard” motor adjustment, reduced fuel consumption goes with an increase in noise production. Therefore, engineers in the automotive industry plan to implement “hard” motor adjustments because of the reduction in fuel consumption and to reduce at the same time the problems resulting in sound quality by absorption. In this context, data displayed in Figure 19 show the ranking of sound quality for the sound of a Diesel motor which had been modified by simulated absorptive measures (Patsouras et al. 2001).

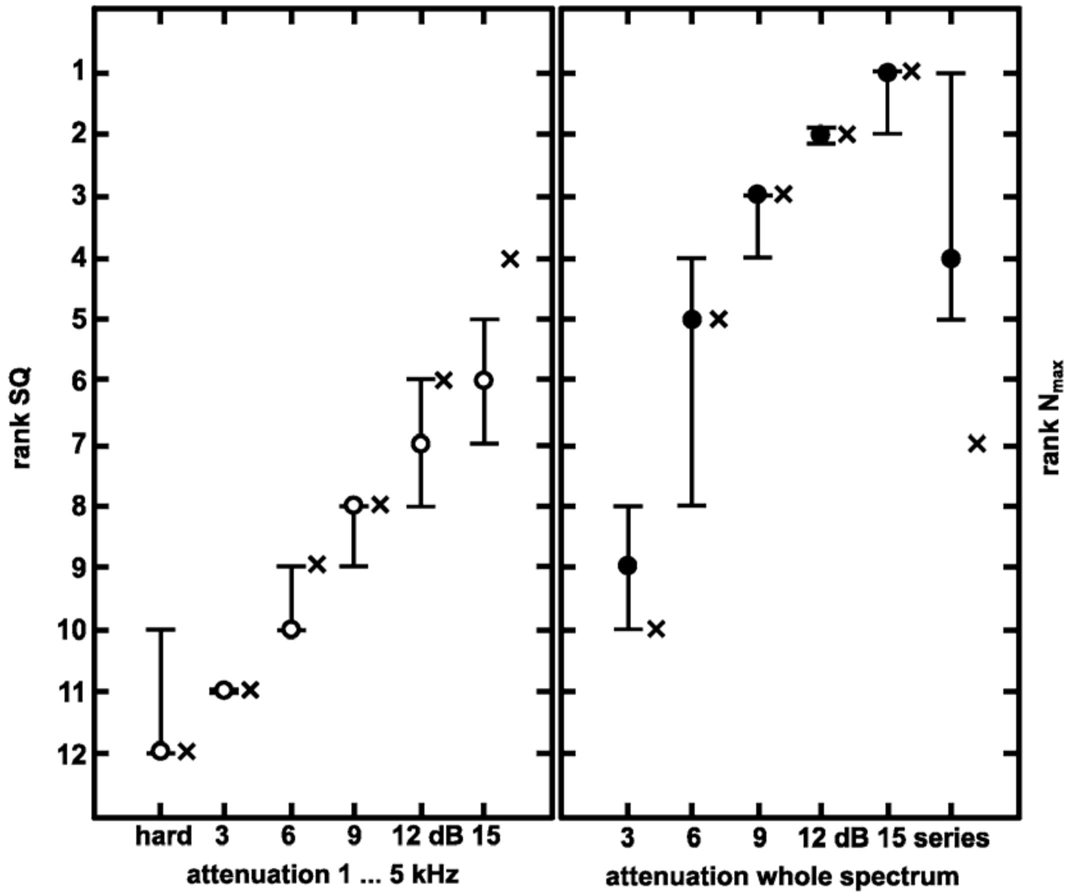

Figure 19: Improvement of the sound quality of a Diesel motor with “hard” motor adjustment when adding simulated absorptive measures. Circles: subjective sound quality estimates, crosses: physical measurements of loudness (after Patsouras et al. 2001)

The data plotted in Figure 19 illustrate the ranking of sound quality of a Diesel motor with “hard” motor adjustment when absorptive measures between 1 and 5 kHz (left panel) or of the whole spectrum (right panel) are simulated. The best ranking (1) is obtained when the whole spectrum is attenuated by 15 dB. The worst ranking (12) of course gets the original “hard” Diesel motor without any absorptive measures. When comparing the subjective evaluations (circles) with physical measurements (crosses), frequently an agreement can be seen within the inter-quartile ranges. However, for a motor which is used in present series vehicles, the subjective evaluation indicates rank 4 whereas the physical evaluation indicates rank 7. On the contrary, an attenuation of 15 dB in the frequency region between 1 and 5 kHz leads to rank 6 in the subjective evaluation, but to rank 4 in the physical evaluation. These discrepancies may be partly due to the fact that the sound of the series motor is well known to the subjects and therefore gets a better ranking. On the other hand, the motor for which the high frequency components are attenuated by as much as 15 dB produces a rather artificial sound, which gets a worse ranking.

3.3 Combined metrics for sound quality evaluation

When assessing sound quality of products usually combined metrics are used. In many cases, a combination of loudness and sharpness can describe main effects of sound quality. In the following, two rather different examples are given: In cooperation with colleagues from the ENT clinic of the Ludwig-Maximilians-Universität München, snoring sounds of different origin and different intensity were recorded (Dreher et al. 2004). In psychoacoustic experiments, their relative annoyance as well as their loudness and sharpness were evaluated. Results are displayed in Figure 20.

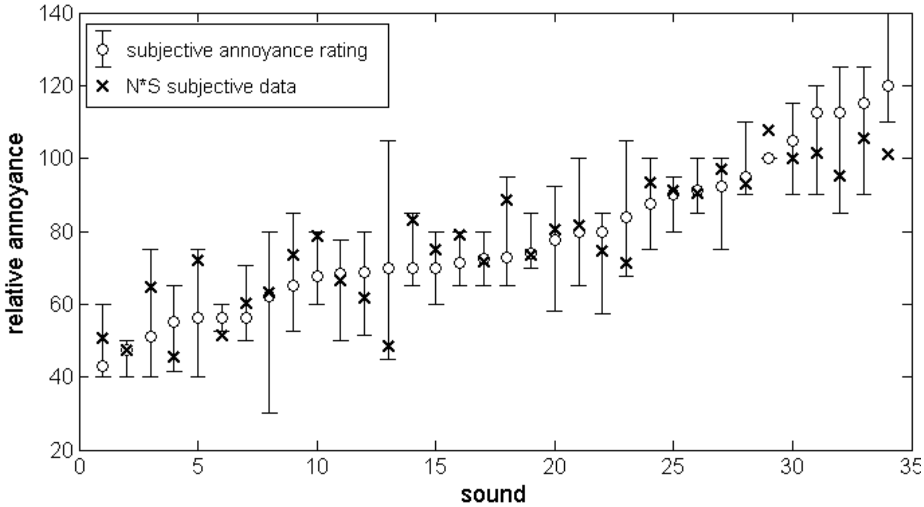

Figure 20: Relative annoyance of snoring sounds. Circles: subjective evaluation of annoyance (medians and inter-quartiles), crosses: relative annoyance calculated from the subjective evaluations of loudness N and sharpness S

The data plotted in Figure 20 indicate that the annoyance rating of snoring sounds can show large inter-individual differences (e.g. sound 8, sound 13). Generally speaking, however, the annoyance of the snoring sounds can be predicted within the inter-quartiles by the evaluation of loudness and sharpness.

Another example comes from the sound quality rating of grand pianos. In psychoacoustic experiments, Valenzuela (1998) evaluated the sound quality of grand pianos from Bösendorfer, Ibach, Steinway, and Yamaha. In addition, Valenzuela proposed a metric, based essentially on physical measurements of loudness and sharpness, which can account for the sound quality rating of the grand pianos.

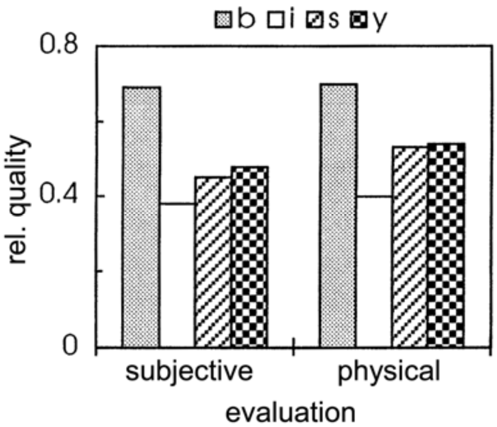

Figure 21: Rating of relative sound quality of four grand pianos from different manufacturers. Subjective evaluation (left), and physical evaluation (right) after Valenzuela (1998).

The data displayed in Figure 21 clearly reveal that the rating of the sound quality of grand pianos can be predicted on the basis of physical measurements of loudness and sharpness. More specifically it should be stated that in addition to sharpness, a weighted distribution of specific loudness with a maximum at 1 kHz is used. For details the reader is referred to the original work of Valenzuela (1998).

In the literature, two different combined metrics for sound quality evaluation have been proposed, which address different practical applications: The concept of sensory pleasantness put forward by Terhardt can assess besides technical sounds also singing voices and musical instruments (Terhardt and Stoll 1978). The concept is illustrated in Figure 22.

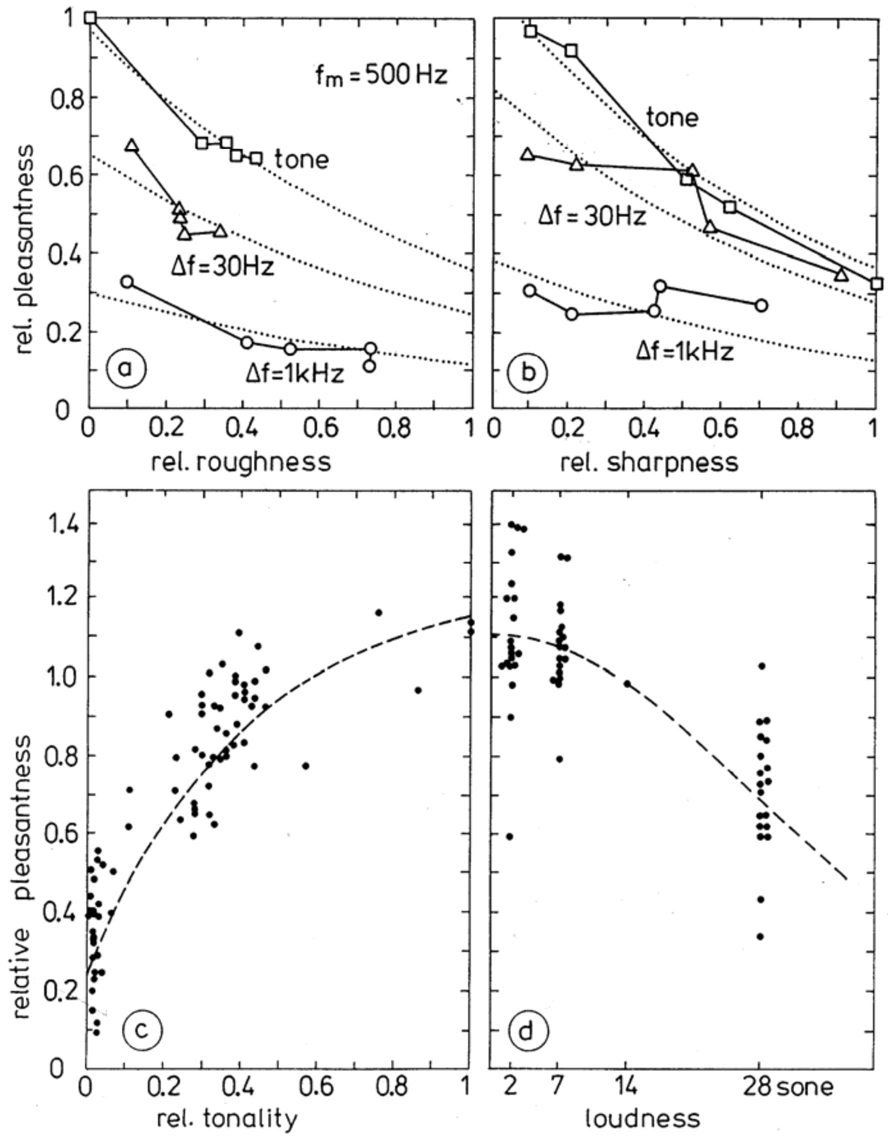

Figure 22: Concept of sensory pleasantness according to Terhardt and Stoll (1978).

The relative sensory pleasantness *decreases* with increasing roughness, with increasing sharpness, and increasing loudness. However, sensory pleasantness *increases* with increasing tonality or pitch strength. This means that in the concept of sensory pleasantness, tonal sounds are rated higher than sounds of stochastic nature. Therefore, sensory pleasantness is very successful in rating musical sounds, where it can be expected that sensory pleasantness increases with the pitch strength. On the other hand, for industrial noise, tonal components - in particular if they are perceived over long periods of time and show slight frequency deviations - can be pretty annoying. Therefore, it is not recommended to use the concept of sensory pleasantness for the description of noise immissions.

For technical sounds, Widmann (1998) proposed a combined metric which can be used for noise emissions as well as noise immissions. The ingredients are illustrated in Figure 23.

$$PA = N_5 \left(1 + \sqrt{w_S^2 + w_{FR}^2}\right)$$

with

– N_5 percentile loudness in sone

– $$w_S = \left(\frac{S}{\text{acum}} - 1.75\right) \cdot 0.25 \lg \left(\frac{N_5}{\text{sone}} + 10\right) \quad \text{for} \quad S > 1.75\,\text{acum}$$

describing the effects of sharpness S and

– $$w_{\text{FR}} = \frac{2.18}{(N_5/\text{sone})^{0.4}} \left(0.4 \cdot \frac{F}{\text{vacil}} + 0.6 \cdot \frac{R}{\text{asper}}\right)$$

describing the influence of fluctuation strength F and roughness R.

Figure 23: Concept of psychoacoustic annoyance proposed by Widmann (1998).

As shown in Figure 23, psychoacoustic annoyance PA depends crucially on the percentile loudness N_5, representing that specific value of loudness which is reached or exceeded in 5 % of the measurement time. In addition to loudness also sharpness S as well as fluctuation strength F and roughness R play an important part (cf. Fastl and Zwicker 2007). The details, how these hearing sensations are included in the calculation of psychoacoustic annoyance are given in Figure 23.

From a more perceptual point of view it shall be stated that psychoacoustic annoyance is strongly influenced by loudness and sharpness as was shown previously for largely different sounds. While loudness describes the power and volume of a sound, sharpness is correlated mainly to the tone color. For both hearing sensations, simulations are at hand which are standardized in DIN 45 631 for loudness and DIN 45 692 for sharpness. The remaining hearing sensations fluctuation strength and roughness deal with temporal variations of sounds. While fluctuation strength describes slower variations (a few per second), roughness deals with sound variations of higher speed (cf. Fastl and Zwicker 2007). The maximum of fluctuation strength is obtained at a modulation frequency of about 4 Hz, the maximum of roughness at about 70 Hz modulation frequency. For temporal variations beyond 200 to 500 Hz, their influence on sound quality evaluation vanishes.

3.4 Sound quality engineering

Not only the evaluation of already existing sounds can be assessed by psychoacoustic metrics (e.g. Mellert et al. 1988), but also desired sounds can be “tailored” using knowledge of psychoacoustics.

Figure 24 shows a typical example for the engineering of warning signals.

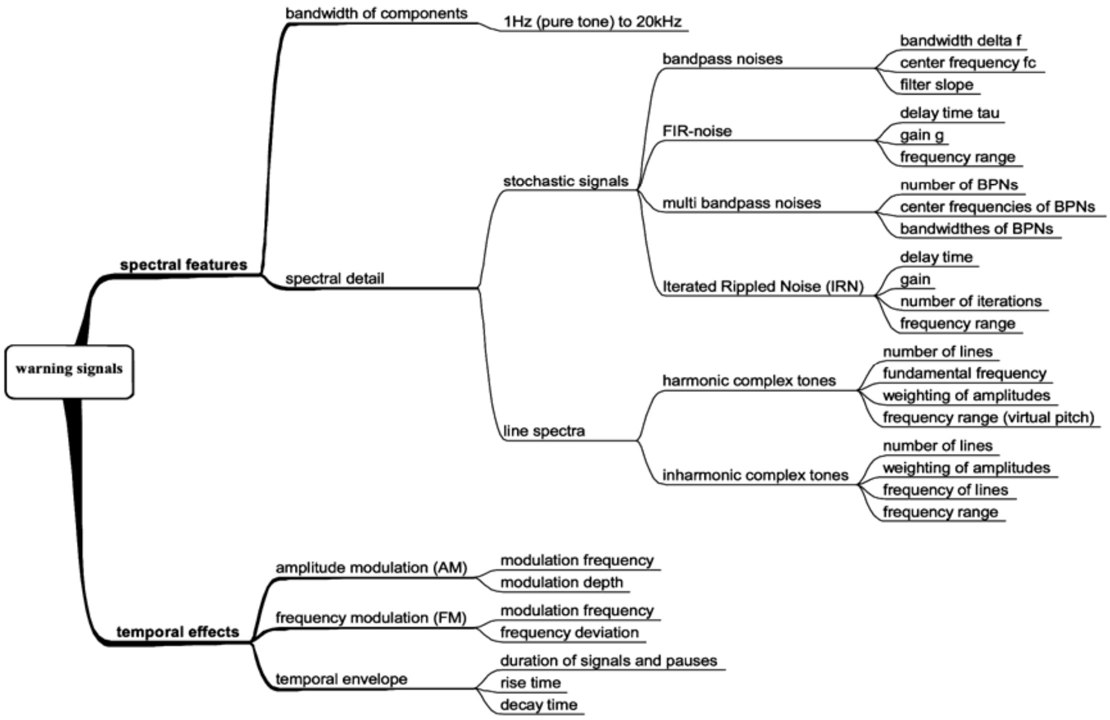

Figure 24: Sound engineering of warning signals. Possible alternatives are treated in a manner typical for decision trees.

When developing a warning signal, several ingredients are taken into consideration. Usually, the sound designer differentiates spectral features and temporal features. With respect to the spectral features it is possible to use just pure tones or a mixture of tones. In both cases, the signals can be characterized by line spectra. On the other hand, it is possible to use stochastic signals for sound quality design. In Figure 24 typical examples are mentioned like band-pass noises, FIR noises, iterated ripple noises, and so forth.

As concerns the temporal effects, amplitude modulation as well as frequency modulation is used. In addition, the warning signals usually get a repetitive temporal envelope. In many cases, repetition rates typical for fluctuation strength can be successfully applied in the design of warning signals. Most important for the sound engineer is the intended use of the warning signal. On the one, hand such a signal just can give some information to the operator. On the other hand, however, it can convey an alert that a life-threatening danger is ahead. Depending on the purpose of the signal, different strategies have to be followed in sound quality engineering.

This approach is not confined to warning signals, but in almost any case of sound quality design, the intended purpose has to be defined beforehand (e.g. Blauert 1986, Jekosch 1999). From the automotive industry an example is given from the work of Zeitler et al. (2006) with respect to the character of different types of vehicles.

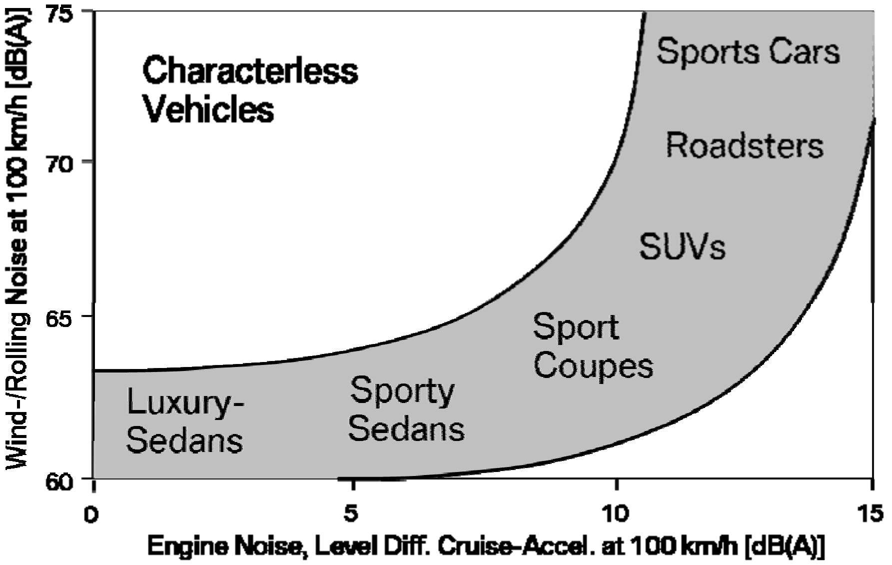

Figure 25: Acoustic vehicle positioning of a manufacturer of premium cars. Different sound characters of the products as defined by the noise levels during cruise (y-axis) or acceleration (x-axis) (from Zeitler et al. 2006).

The data displayed in Figure 25 clearly indicate that different types of cars go with rather different target sounds. For example, luxury sedans should have low noise during cruising and little level difference correlated with acceleration. On the other hand, for sports cars higher wind and rolling noises are typical, and the driver wants not only to feel but also to hear the acceleration by a considerable increase of the sound with acceleration.

4 MUSIC

4.1 Musical consonance and roughness

As mentioned already earlier, music has strong ties towards psychoacoustics and sound quality engineering. These ties shall be elaborated first by the example of roughness. Almost 150 years ago, Helmholtz (1863) based his concept of musical consonance on the absence of the hearing sensation roughness. His arguments are illustrated by means of Figure 26.

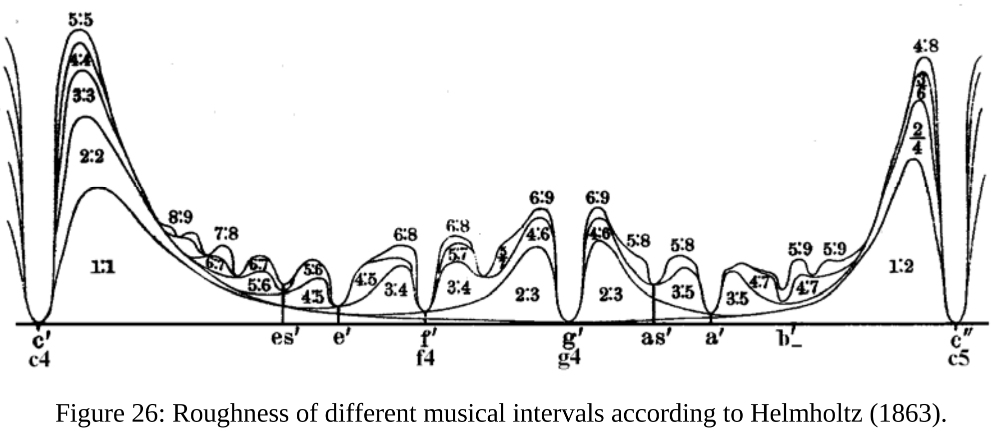

Figure 26: Roughness of different musical intervals according to Helmholtz (1863).

When starting from C_4, the octave C_5 represents a perfect consonance which also leads to no roughness. Likewise the fifth G_4 as a consonant interval also is free of roughness. More or less the same holds true for the fourth F_4. On the contrary, the interval of a semitone (C_4#) produces large values of roughness as does the interval C_4 - B_4.

Interestingly, the concept of roughness as put forward by Helmholtz (1863) and formulated in modern terms by Terhardt (1968), originally was treated in a musical context, namely to distinguish consonant and dissonant musical intervals. However, these days, roughness is one of the important ingredients in sound quality design. For example, in the automotive industry it is well known that the sound of a sportive car has to include just the right amount of roughness. If, however, too much roughness is added to the car sound, it can resemble the noise produced by a tractor, eliciting the inverse image from a sportive car.

4.2 Musical dynamics and loudness scaling

Another feature of music which is frequently applied both in psychoacoustics and sound quality rating is the representation of musical dynamics. It is well known from music theory that musical dynamics ranges between very soft (pianissimo, pp) and very loud (fortissimo, ff). For even more extreme values of loudness, in the musical notations piano-pianissimo (ppp) and forte-fortissimo (fff) are used. These indications of musical dynamics with their century old tradition are adapted for rating of loudness in category scales. This reasoning is illustrated in Figure 27.

pp	***p***	***mf***	***f***	***ff***
very soft	soft	neither soft nor loud	loud	very loud

ppp	***pp***	***p***	***mf***	***f***	***ff***	***fff***
extremely soft	very soft	soft	neither soft nor loud	loud	very loud	extremely loud

Figure 27: Musical dynamics and five step category scale (upper panel) or seven step category scale (lower panel).

The data displayed in Figure 27 indicate the good correspondence between the symbols of musical dynamics and loudness rating by category scales. However, for the middle category in musical notation (mezzo-forte, mf), the interpretation in music versus psychoacoustics or sound quality engineering is a little different: In sound quality engineering, the category “neither soft nor loud” is exactly in the middle between the category soft and the category loud. In the musical notation, however, mezzo-forte literally means “half loud”. Therefore it is not centered between the categories “soft” (p) and “loud” (f), but mezzo-forte (mf) is closer to the category “loud” (f).

The same holds true for the seven step category scale displayed in the lower panel of Figure 27. Again “neither soft nor loud” is just between “soft” and “loud”, whereas mezzo-forte is a little closer to “forte” in comparison to “piano”.

Despite these slight differences between musical dynamics and category scaling of loudness in psychoacoustics and sound quality engineering, both five step scales and seven step scales have proven very successful in practical applications of sound quality.

A related aspect of music which is reflected also in psychoacoustics and sound quality rating are sometimes misleading predictions of loudness when using A-weighted level (e.g. Möhler 1988, Vos 2006). An example from music is given in Figure 28.

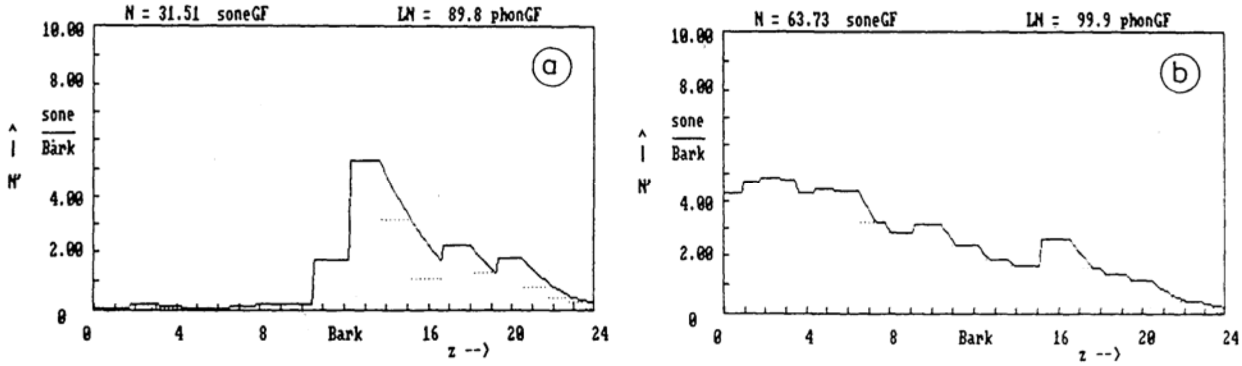

Figure 28: Loudness pattern for a piccolo flute (a) versus a pipe organ with full registers (b) at same A-weighted level of 85 dB(A).

The loudness patterns displayed in Figure 28 clearly indicate the large differences in tone color between a piccolo flute (left) and a pipe organ with full registers (right) at same A-weighted level. Both in musical acoustics and in sound quality engineering, a single channel analysis by A-weighted level can not account for differences in tone color. The loudness patterns clearly reveal that the piccolo flute produces mainly high frequency sounds in excess of 13 Bark (2 kHz). On the other hand, the right panel in Figure 28 indicates that a pipe organ with full registers produces sounds in the whole range of hearing. It is well known that the area of the loudness patterns is directly proportional to the perceived loudness. This means that at same A-weighted level of 85 dB(A), the pipe organ at full registers produces about twice the loudness in comparison to the piccolo flute.

In order to illustrate for a broad public the loudness of different sounds, level-thermometers are very popular. However, for sounds of different tone color, the indications on the level-thermometer can be misleading. An example is illustrated in Figure 29.

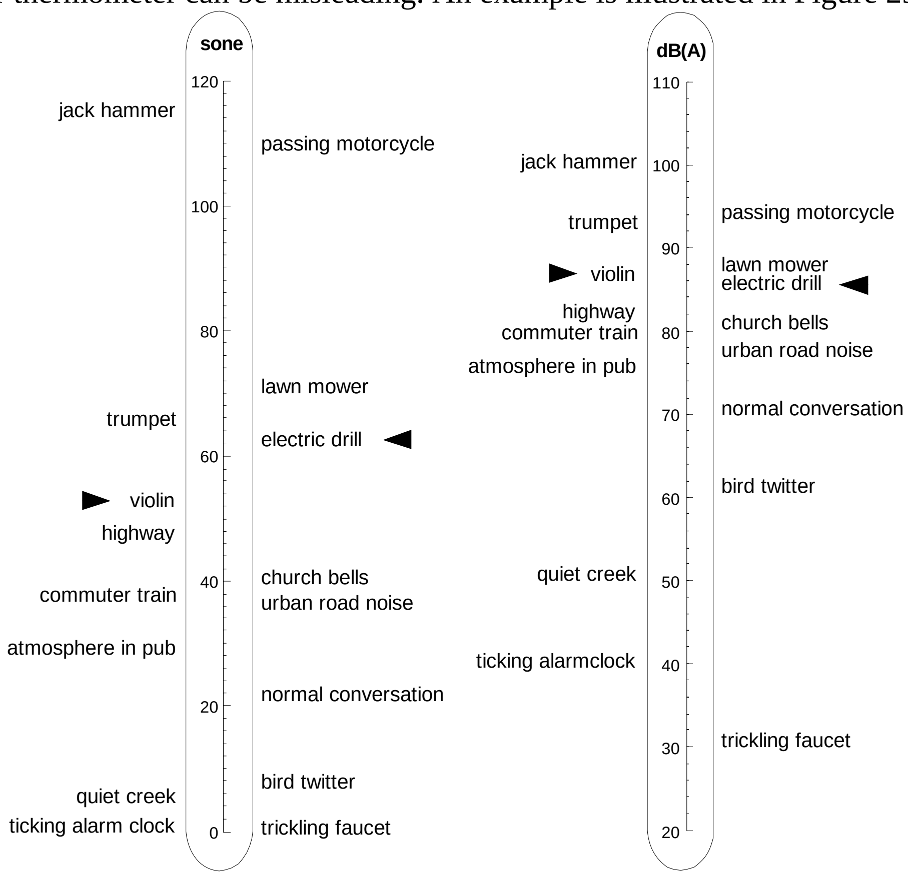

Figure 29: Loudness-thermometer (left) in comparison to a level-thermometer (right), from Fastl et al. 2006

The level-thermometer shown in the right part of Figure 29 indicates that the sound of a violin or the sound of an electric drill produce the same A-weighted level. However, the loudness-thermometer in the left part of Figure 29 illustrates that the loudness of the electric drill is significantly larger than the loudness of the violin sound.

Another pair of sounds shows the opposite behavior: In the loudness-thermometer (left), the sound of a trumpet produces less loudness compared to the sound of a lawn mower. However, in A-weighted level, as illustrated in the right part of Figure 29, the trumpet produces a higher A-weighted level than the lawn mower. In conclusion then, a loudness-thermometer is preferred to a level-thermometer because it rates the loudness of different product sounds in line with the subjectively perceived loudness.

Since compared to their A-weighted level, musical instruments are considered to produce smaller loudness values (cf. Figure 29) one might suspect that also cognitive effects may play a role. A corresponding hypothesis would go as follows: Since the sounds of musical instruments are preferred in comparison to technical sounds, the loudness rating of musical sounds would be lower.

In order to check the validity of such a hypothesis, a procedure was developed (Fastl 2001) which largely obscures the information about the sound source. The block diagram illustrating the procedure is given in Figure 30.

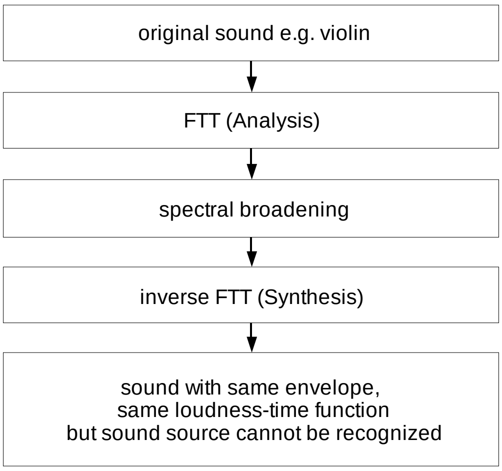

Figure 30: Block diagram illustrating the procedure to obscure the information about the sound source of a stimulus (after Fastl 2001).

In the procedure illustrated in Figure 30, the original sound first is analyzed by Fourier-Time-Transform (FTT, Terhardt 1985) and after spectral broadening re-synthesized by inverse FTT. In this way, a sound with the same envelope and the same loudness-time function is obtained, but the sound source can no longer be recognized.

Figure 31 illustrates by an example the related signal processing.

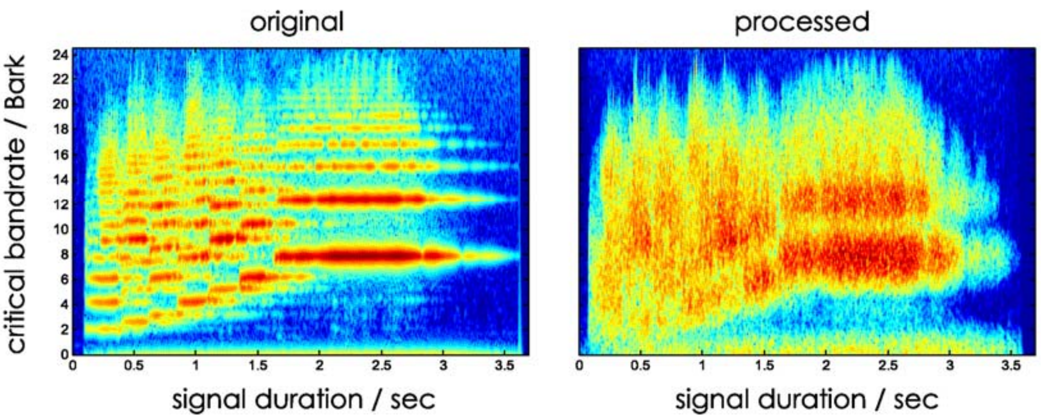

Figure 31: Example for obscuring the information about the sound source. FTT-spectrum of the original sound (left) and the related FTT-spectrum of the processed sound (right) in which the information about the sound source is obscured.

The FTT-spectrum in the left part of Figure 31 shows that a scale is played on a violin. The harmonic structure of the violin sound as well as the increasing pitch of the notes on the scale are clearly visible. In the right part of Figure 31, the increase in pitch can also be detected, and there is some indication of a harmonic spectral structure. However, in comparison to the FTT-spectrum of the original sound (left), the FTT-spectrum of the processed sound is blurred and all detail is lost.

Extended experiments with original and processed sounds (Fastl et al. 2006) revealed that loudness evaluation is hardly influenced by the recognition of the sound source. These results are in line with data of Ellermeier et al. 2004a, Hellbrück et al. 2004, as well as Zeitler et al. 2004. On the other hand, however, cognitive effects can play an important role in annoyance rating (e.g. Ellermeier et al. 2004b). In this case, the recognition of the sound source (e.g. clinking wine glasses) can lead to a substantially lower annoyance rating than for the processed version, despite almost identical loudness-time functions.

5 OUTLOOK

In this paper, relations between psychoacoustics, sound quality and music are discussed. Several of the features studied for centuries in musical acoustics lead to very helpful tools in both psychoacoustics and sound quality engineering.

The Dynamic Loudness Model (DLM) is getting more and more accepted in practical applications, because it can assess loudness evaluation also by persons with (slight) hearing deficits, which – unfortunately - these days are very common in industrialized countries. The concept of pitch strength can be used for noise immissions from large industrial plants as well as components of a car. Since in a modern premium class car, more than hundred electric motors can be found, the rating of possible whine and other undesired tonal components by pitch strength is highly recommended. For sound quality rating, in many cases the main aspects are loudness and sharpness. Recently new draft standards for loudness of temporally varying sounds (DIN 45 631) as well as sharpness (DIN 45 692) have been published which strongly support the application of these psychoacoustic magnitudes. Moreover, it is to be expected that in addition to physical and psychoacoustic aspects, cognitive effects will increase in relevance with respect to sound quality evaluation and sound quality design.

6 ACKNOWLEDGEMENTS

The author wishes to thank former and present members of his group AG Technische Akustik, MMK, TU München for numerous contributions to the work presented here. Dipl.-Ing. Florian Völk is acknowledged for substantial editorial help. Large parts of the research have been supported over years by the Deutsche Forschungsgemeinschaft.

7 REFERENCES

[1] Jont B. Allen, “Harvey Fletcher's role in the creation of communication acoustics,” Journal of the Acoustical Society of America, **99**(4), 1825-1839 (1996).

[2] Faruk Bayraktar, H. Temel Belek, “Prediction of the side panel vibrations on the washing machines by the elastic multi-body modeling,” *Proceedings of INTER-NOISE 06* (2006).

[3] Jens Blauert, “Cognitive and esthetic aspects of noise engineering,” *Proceedings of INTER-NOISE 86*, vol I, 5-14 (1986).

[4] Josef Chalupper, *Perzeptive Folgen von Innenohrschwerhörigkeit: Modellierung, Simulation und Rehabilitation.* (Shaker, Aachen, 2002).

[5] Josef Chalupper and Hugo Fastl, “Dynamic loudness model (DLM) for normal and hearing-impaired listeners,” Acta Acustica united with Acustica, **88**, 378-386 (2002).

[6] DIN 45 631, *Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum, Verfahren nach E. Zwicker.* (Deutsches Institut für Normung e. V., Berlin, 1991).

[7] DIN 45 692, *Messtechnische Simulation der Hörempfindung Schärfe*. (Deutsches Institut für Normung e. V., Berlin, 2007).

[8] Alfred Dreher, Richard de la Chaux, Christine Klemens, Andrea Nobbe, Uwe Baumann, Gerd Rasp, “Frequenzanalyse und pharyngeale Druckmessung in der Topodiagnostik des Schnarchens,” Laryngo-Rhino-Otologie, **83**(23-3) (2004).

[9] Wolfgang Ellermeier, Alfred Zeitler, Hugo Fastl, “Impact of source identifiability on perceived loudness,” *Proceedings of the 18. ICA Kyoto*, 1491-1494 (2004a).

[10] Wolfgang Ellermeier, Alfred Zeitler, Hugo Fastl, “Predicting annoyance judgments from psychoacoustic metrics: Identifiable versus neutralized sounds,” *Proceedings of INTER-NOISE 04* (2004b).

[11] Hugo Fastl, “Pitch strength of pure tones,” *Proceedings 13. ICA Belgrade*, Vol. 3, 11-14 (1989).

[12] Hugo Fastl, “The Psychoacoustics of Sound-Quality Evaluation,” ACUSTICA - acta acustica **83**, 754-764 (1997).

[13] Hugo Fastl, “Neutralizing the meaning of sound for sound quality evaluations,” *Proceedings of the 17th International Congress on Acoustics* (2001).

[14] Hugo Fastl and Gerhard Stoll, “Scaling of pitch strength,“ Hearing Research, **1**, 293-301 (1979).

[15] Hugo Fastl and Helmut Fleischer, “Über die Ausgeprägtheit der Tonhöhe von Paukenklängen,“ *Fortschritte der Akustik, DAGA’92*, 237-240 (DPG GmbH, Bad Honnef, 1992).

[16] Hugo Fastl and Eberhard Zwicker, *Psychoacoustics – Facts and Models.* (3rd edition, Springer, Heidelberg New York , 2007).

[17] Hugo Fastl, Daniel Menzel, Michael Krause, “Loudness-thermometer: evidence for cognitive effects?” *Proceedings of INTER-NOISE 06* (2006).

[18] Markus Fruhmann, “Ein Modell der Ausgeprägtheit der Tonhöhe in Theorie und Praxis,“ *Fortschritte der Akustik, DAGA ’06*, 315-316 (DEGA, Berlin, 2006a).

[19] Markus Fruhmann, *Ein Modell zur Wahrnehmung der Ausgeprägtheit der Tonhöhe*. (PhD Thesis, Technische Universität München, 2006b).

[20] Takeo Hashimoto, “Sound quality approach on vehicle interior and exterior noise: Quantification of frequency related attributes and impulsiveness,” Journal of the Acoustical Society of Japan (E)**21**, 337-340 (2000).

[21] Aaron Hastings and Patricia Davies, “An exploration of the relationship between annoyance ratings and temporal variations in diesel engine noise,” *Proceedings of INTER-NOISE 06* (2006).
[22] Jürgen Hellbrück and Wolfgang Ellermeier, *Hören. Physiologie, Psychologie und Pathologie.* (2. Auflage, Hogrefe, Göttingen, 2004).
[23] Jürgen Hellbrück, Hugo Fastl, Barbara Keller, “Does Meaning of Sound influence Loudness Judgements?” *Proceedings of the 18. ICA*, 1097-1100 (2004).
[24] Hermann von Helmholtz, *Die Lehre von den Tonempfindungen.* (Vieweg, Braunschweig, 1863).
[25] ISO 532 *Acoustics -- Method for calculating loudness level* (ISO, International Organisation for Standardization, 1975).
[26] Ute Jekosch. “Meaning in the context of sound quality evaluation,” Acta Acustica, **85**, 681-684 (1999).
[27] Birger Kollmeier, *Psychoacoustics, Speech and Hearing Aids.* (World Scientific, Singapore, 1996).
[28] Sonoko Kuwano, Hugo Fastl, Seiichiro Namba, Seishi Nakamura, Hiroshi Uchida, ”Quality of Door Sounds of Passenger Cars,” *18. ICA Kyoto*, 1365-1368 (2004).
[29] Sonoko Kuwano, Hugo Fastl, Seiichiro Namba, Seishi Nakamura, Hiroshi Uchida, ”Quality of Door Sounds of Passenger Cars,” Acoustical Science and Technology, **27**(5), 309-312 (2006).
[30] Volker Mellert, Reinhard Weber, Klaus Betke, Hermann Remmers, “Noise Source Identification and Machinery Diagnosis by Ear-Related Acoustic Parameters,” *Proceedings of INTER-NOISE 88*, 79 (1988).
[31] Ulrich Möhler, “Community response to railway noise: a review of social surveys,” Journal of Sound and Vibration, **120**, 321-332 (1988).
[32] Christine Patsouras, *Geräuschqualität von Fahrzeugen - Beurteilung, Gestaltung und multimodale Einflüsse*. (Shaker , Aachen, 2003).
[33] Christine Patsouras, Hugo Fastl, Dimitrios Patsouras, Klaus Pfaffelhuber, “Subjective Evaluation of Loudness Reduction and Sound Quality Ratings Obtained with Simulations of Acoustic Materials for Noise Control,” *Proceedings Euro Noise 2001*, CD-ROM (EAA, 2001).
[34] Christine Patsouras, Hugo Fastl, Ulrich Widmann, Georg Hölzl, “Psychoacoustic evaluation of tonal components in view of sound quality design for high speed train indoor noise,” Acoustical Science and Technology, **23**, 113-116 (2002).
[35] Ernst Terhardt, “Über akustische Rauhigkeit und Schwankungsstärke,” Acustica, **20**, 215-224 (1968).
[36] Ernst Terhardt, “Fourier transformation of time signals: Conceptual revision,” Acustica, **57**, 242-256 (1985).
[37] Ernst Terhardt and Gerhard Stoll, “Bewertung des Wohlklangs verschiedener Schalle,“ *Fortschritte der Akustik, DAGA’78*, 583-586 (VDE-Verlag, Berlin, 1978).
[38] Miriam N. Valenzuela, *Untersuchungen und Berechnungsverfahren zur Klangqualität von Klaviertönen.* (Herbert Utz Verlag Wissenschaft, München, 1998).
[39] Florian Völk, Hugo Fastl, Markus Fruhmann, Stefan Kerber, “Psychoakustische Untersuchungen zu Inharmonizitäten von Gitarrensaiten,“ *Fortschritte der Akustik, DAGA ’06*, 743-744 (DEGA, Berlin, 2006).
[40] Joos Vos, “Noise annoyance caused by mopeds and other traffic sources,” *Proceedings of INTER-NOISE 06* (2006).
[41] Ulrich Widmann, “Aurally adequate evaluation of sounds,” *Proceedings of Euro Noise 1998*, edited by Hugo Fastl and Joachim Scheuren, 29-46 (DEGA, Oldenburg, 1998).

[42] Norbert Wiesmann and Hugo Fastl, “Ausgeprägtheit der Tonhöhe und Frequenzunterschiedsschwellen von Bandpass-Rauschen,“, *Fortschritte der Akustik, DAGA’91*, 505-508 (DPG GmbH, Bad Honnef, 1991).
[43] Alfred Zeitler, Wolfgang Ellermeier, Hugo Fastl, “Significance of Meaning in Sound Quality Evaluation,” *Fortschritte der Akustik, CFA/DAGA 04*, 781-782 (2004).
[44] Alfred Zeitler, Jürgen Hellbrück, Wolfgang Ellermeier, Hugo Fastl, Gerhard Thoma, Peter Zeller, “Methodological approaches to investigate the effects of meaning, expectations and context in listening experiments,” *Proceedings of INTER-NOISE 06* (2006).
[45] Manfred Zollner, *Die Physik der Elektrogitarre*. (preprint 2007).
[46] Eberhard Zwicker, “Ein Verfahren zur Berechnung der Lautstärke,” Acustica, **10**, 304-308 (1960).