Influence of context effects on sound quality assessments

André Fiebig
HEAD acoustics GmbH, Herzogenrath, Germany

Summary
The process of (product) sound quality assessment usually takes place in complex settings. Accordingly, in addition to the perception of sound character representing basic auditory sensations further information is processed finally resulting in a sound quality assessment.
According to Blauert a quality judgment starts out from a set of recognized features compared to a "reference set" of features and the distance between these two sets represents the perceived quality. Spatial, time, semantic, multimodal or response context substantially influence the perceived distance between recognized and expected features. Moreover, auditory sensations, expected to represent lower cognitive processes, are prone to context effects as well. But, the influence of context effects on sound perception should not be understood as a kind of bias, but such phenomena are related to the way humans perceive their environment in everyday life.
The paper illustrates the importance of the sound character and the impact of contextual circumstances on the assessment of sound quality by means of case studies. It is intended to provide a deeper understanding of modifying factors beyond the acoustical stimulus, which are integrated into the (quality) judgment of sound.

PACS no. 43.50.Ba, 43.50.Qp

1. Introduction

It is broadly acknowledged that several non-acoustical aspects exert influence on the perception and assessment of sound. Psychophysical tasks are rich in psychological processes and decisions depend on memory, on comparative behavior, and on response strategies. This applies particularly to sound quality assessments, where recognized sound features are compared to a reference set of features in order to derive a sound quality impression [1]. For example, House and Shively reported on contextual effects with regard to sound quality assessment of loudspeakers. They observed that sound quality judgments were significantly biased by non-auditory factors such as size, price and brand name [2]. Thus, since the perception and assessment of sound quality is context-dependent, the quality of an acoustical stimulus referring to sound quality cannot be determined without addressing contextual variables like spatial, time, semantic or response context [3].
Therefore, it is important to study the influence of non-acoustical aspects on sound quality assessments more in detail to be able to design sound quality appropriately ensuring high sound quality assessments in everyday life applications.

1.1 Meaning of Context

Context effects and cognitive biases refer to effects, which are related to patterns of deviation in (sound) judgments. For example, anchoring effects, contextual biases, test setting effects or priming are frequently observed in listening experiments and are described as cognitive bias effects. It is acknowledged that a stimulus context can affect processes occurring at every stage, from early sensory transduction, perceptual encoding to cognitive recoding and decision [4]. This means that if context effects are disregarded, the comparability of sound perceptions (responses) is limited. It is important to point out that any bias due to context effects should not be interpreted as undesirable, since such biases present exactly the nature of human (multimodal) experience. According to Haselton, it is important to know the nature and extent of biases not only with respect to the validity of experimental results, but also because these influences reveal the design of the mind [5]. A lack of awareness of biases reduces the explanatory power of experimental results. Context effects refer to systematic distortions beyond occasional and accidental errors. Thus, such influences can be interpreted as systematic deviations from "expected" responses.

1.2 Sound character vs. sound quality

The terms sound character and sound quality were introduced to distinguish between different phenomena. The term sound character refers to the perception related to basic attributes of auditory events without considering context, action and higher level of cognitive processing. This means that it is not colored by any context. In contrast to sound character, the term sound quality is related to the perceived adequacy of a (product) sound. The perception of adequacy and suitability is affected by diverse factors such as context, cognition, interaction [6].

The sensation of sound character is ideally devoid of any contextual influences and according to the sound character concept is open to a parametric representation of a sound [7]. In contrast to it, sound quality cannot be judged without a reference to a concept of expected and desired features of the considered object. This notion implicates an act of interpretation, which besides the acoustical stimulus itself refers back to the meaning, context and role of the product in the habitat of the evaluator. Although the distinction between sound character and sound quality is helpful, it must be noted that any strict separation of these terms seems incorrect. Sound character, as a perceptual dimension, is defined as being composed of basic auditory sensations without considering context or action. However, any laboratory setup provides a context in itself and therefore, the pure perception of sound character cannot be observed.

2. Sound Quality Case Studies

2.1 Assessment of vacuum cleaner sounds in a home-environment and in laboratory

In a between-subjects design of experiment the influence of different experimental contexts on different vacuum cleaner sounds were investigated. The test setting was changed: the test subjects had to complete after listening to a vacuum cleaner sound a semantic differential in a laboratory as well as in a close to reality context (living room atmosphere). The influence of the test environment on sound assessments was studied by means of a semantic differential test. In the close to reality context a typical additional sound source was present to enhance the naturalness of the test situation. The test subjects had the possibility to watch TV. However, they were requested to judge the sound of different vacuum cleaners only.

2.1.1 Method

Subjects

A semantic differential test was performed by 15 test subjects (9 male, 6 female) in a laboratory context. Moreover, 15 test individuals (10 male, 5 female) participated in a listening experiment, which was performed in a test setting resembling a living room.

Apparatus

Vacuum cleaner product sounds presented in the listening test performed in laboratory context were binaurally measured under controlled measurement conditions in a living room environment.

In the living room situation, the test subjects were exposed to the real vacuum cleaner products emitting the sounds to be judged. In contrast to the laboratory context, in the living-room environment the test subjects were requested to watch TV while the vacuum cleaner products were used by another subject. The TV stimulus was constant for all test subjects (daily news). The test subjects could control the volume of the TV by themselves. All instructions were presented on a computer screen.

Stimuli

Three vacuum cleaner sounds were considered. In living room context, the test subjects listened to the original products emitting the noise, whereas in laboratory context all test subjects were exposed to binaural measurements of the respective vacuum cleaner products used in the first part of the experiment (see figure 1). It can be seen that the noises possess a similar time structure; first the device is switched on, then a stationary period of approx. 30 s follows and finally the vacuum cleaner is switched of. Figure 1 illustrates that the sounds differ considerably in sound pressure level and spectrum as well as in the self-cleaning sound after turning off the vacuum cleaner.

The sound exposure situation (distance between source and receiver, operating condition (brush on, moving)) was comparable in both test conditions. The binaural measurements, realized with a calibrated head measurement system HMS IV, were carried out in the living room context.

In general, it has to be remarked that in the living room experiment an additional noise source, a TV, was present. The aim of the study was to investigate the comparability of product sound quality assessments between laboratory context, where all confounding variables are kept constant, and a more realistic test setting, where further noise sources are usually present besides the product under scrutiny.

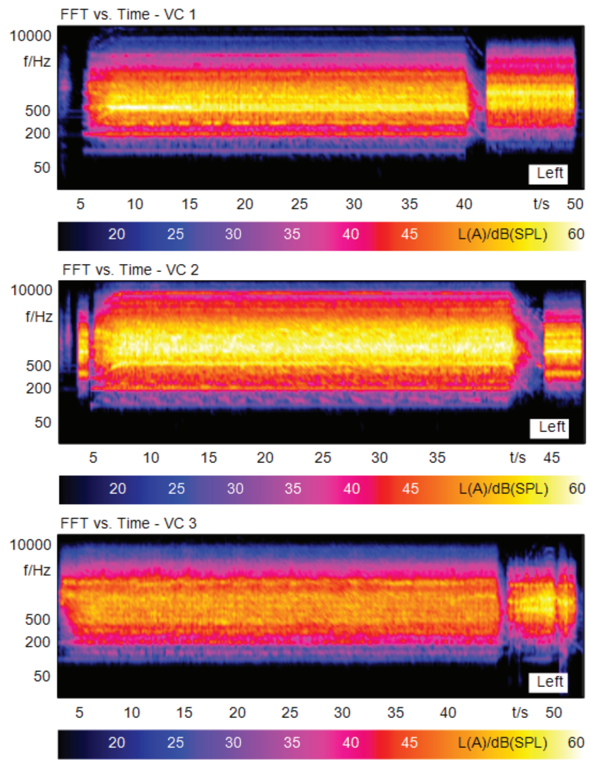

Figure 1. FFT vs. time of three vacuum cleaner sounds

The test subjects could adjust the volume of the TV to realize a typical living home situation. The volume level change was dependent on the vacuum cleaner used. The test subjects adjusted the volume level of the TV in dependence of the respective vacuum cleaner in a statistically significant different way (*p<0.01***) (see fig. 2). This means that the acoustical stimuli were not identical in both experimental set-ups.

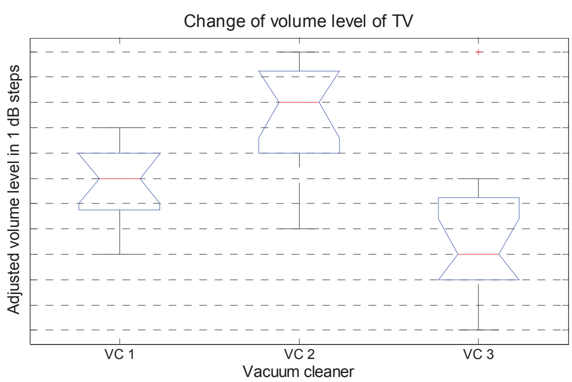

Figure 2. Volume level change of TV stimulus during vacuum cleaner sound presentation in living room context

Procedure

The sounds were judged on eleven 7-point bipolar category scales in a semantic differential test. The evaluation criteria covered different semantic dimensions including valence (pleasantness, harmony), sound character (loud, sharp) and emotion (aggression). A between-subject-design of experiment was applied. In order to provide an associative context comparable to the home environment, the following description was given in laboratory context: "Please imagine, you are sitting in a chair and another person is using a vacuum cleaner." The passive role of the listener in the laboratory context suggested by the instruction is comparable to the passive role of the subject in the living room test context. The test subjects did not mention any difficulties in applying the given associative context as the background for their sound assessments.

2.1.2 Results

The retest-reliability was investigated in laboratory context, since five test subjects performed the full experiment twice with a break of 2 hours between the experiments. The observed retest-reliability was medium (r_{tt}*=0.70***). Possibly due to difficulties in assessing non-auditory related categories, such as reliability, the retest-reliability was not particular high.

As expected, the experimental contexts, laboratory and living room context, cause different judgments and small but consistent differences occur. Figure 3 illustrates exemplarily observed differences between the assessments provided in laboratory context and in living room context.

Although, in several cases the overlapping confidence intervals indicate that the differences are not statistically significant so far, partially considerable differences between the judgments of the same product sound occur. Figure 3 displays that some characteristics were differently judged due to the different test conditions. One-way ANOVA tests confirmed that for some evaluation criteria the effect of test environment on the assessments of product sounds were statistically significant; the evaluation criteria seems to be prone to the effect of test environment. For example, the test environment effect was almost significant for the assessment of power ($F_{(1,84)}=3.8$, $p=0.05$) with a small effect size of $\eta^2=0.04$. In average, test subjects gave 0.5 categories higher assessments of power in laboratory context than in the living room environment, which could be attributed to the focused attention on the acoustical stimulus in laboratory context. The effect of the test environment is significant for the evaluation criteria dull ($F_{(1,84)}=15.4$, $p<0.01$**), sharp

($F_{(1,84)}=8.3$, $p<0.01^{**}$) and almost significant for acceptable ($F_{(1,84)}=5.9$, $p=0.05$). The test subjects judged the vacuum cleaner sounds as more acceptable in living room context. This judgmental behavior might be related to the different level of attention to the sounds in both experimental conditions leading to a changed assessment of few stimuli characteristics.

In summary, the test environment causes already small effects on sound assessments for several evaluations. Interaction effects between vacuum cleaner sound and test environment were not found to be significant, however in few cases an interaction tendency is apparent.

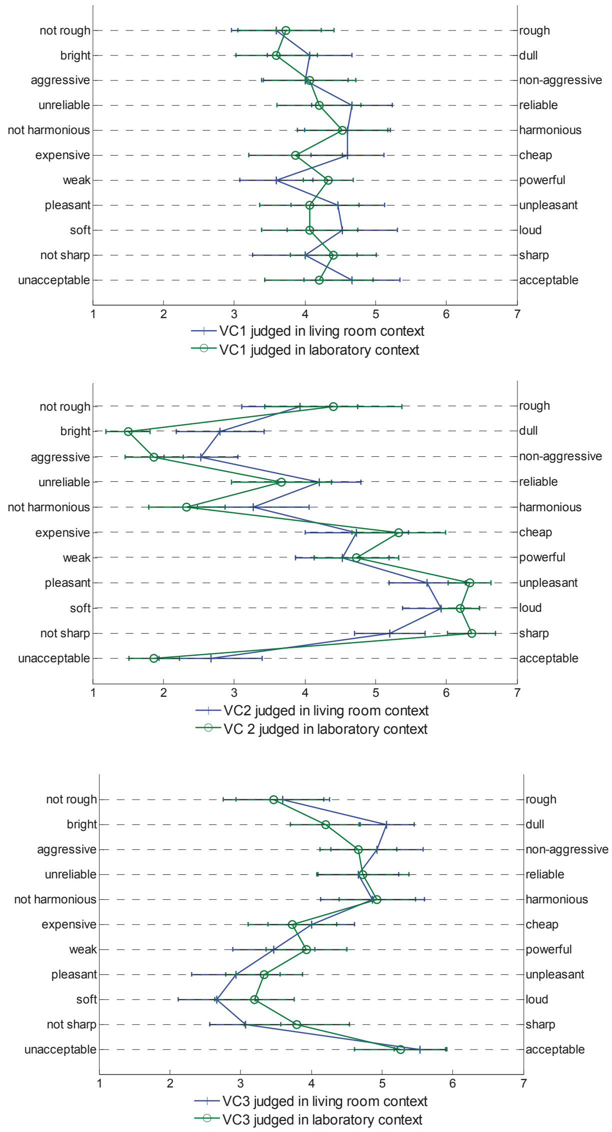

Figure 3. Semantic profile plot: Comparison of sound assessments of three vacuum cleaners peovided in living room and laboratory context (arithmetic mean values and 95 % confidence interval)

2.1.3 Discussion

The comparison between vacuum cleaner sound assessments gained in a classical laboratory experiment and judgments collected in a test design more adapted to daily-life experiences shows an influence of the experimental context on sound assessments. Even if assessments of “meaningless” auditory sensations are requested (like sharpness or loudness) differences explained by the effect of experimental context were observed. Although, several differences are not significant in a statistical sense and the null hypothesis cannot be fully rejected regarding all considered evaluation criteria so far, a trend can be observed. In the living room context, the subjects were less critical about the product sounds and perceived several sound characteristics as less distinct. However, it is important to mention that the test subjects’ assessments were not particularly colored by non-acoustical aspects, like color, design, brand name of the products in the test design more adapted to daily-life experiences, since their attention was intentionally attracted to the TV stimulus. The test subjects mainly based their sound quality assessments on the acoustical stimuli. It is expected that more complex sound assessment processes are triggered, if further product information is processed, which is understood as sound quality perception. Moreover, since the test subjects could adjust the volume of the TV, they probably masked any disturbing vacuum cleaner noise to a certain degree resulting in slightly better sound judgments in the close to reality context. However, the assessments of the vacuum cleaner in the semantic differential experiments were similar. The presence of the “confounding” sound source, the TV, did not lead to different results as well as did not cause greater spread in the data. Significance tests could not determine any significant difference in variance in the close to reality context. This illustrates the ability to judge reliably product sound quality in complex, more realistic test situations.

In general, the systematic judgmental differences are likely related to different levels of attention in the both experimental conditions. This assumption is in line with recent publications [7]. The presented case study demonstrates mainly two aspects: (a) in context of the assessment of product sounds laboratory and close to reality experiments can yield correlated results ($r_{lab/field}=0.87^{**}$), but (b) the congruence of experimental results obtained in different environments is limited. The experiments about the “sound quality” of vacuum cleaner products illustrate that due to the specific experimental context consistent and significant differences can occur, which have to be taken into account with respect to the generality of the experimental results. Although the variance of judgments is mainly influenced by the characteristics of the category scaling of different vacuum cleaner sounds, in the performed case study up to 8 % of the total variance can be accounted for by the factor test environment.

2.2 Assessment of Kettle Noises with/without Visual Information Background of Experiment

In total, the sounds of fourteen different kettle products (water boilers) were subject to sound quality assessments. Five out of the fourteen sounds were presented twice in a laboratory experiment. In a first test session, all fourteen sounds were only acoustically presented. After a break of few minutes, in a second test session in addition to the playback of kettle sounds respective pictures of the different products were shown.

2.2.1 Method

Subjects

Fourteen test subjects (9 male and 5 female) participated in the experiment.

Apparatus

Kettle product sounds were binaurally measured under controlled measurement conditions. The sounds in the experiment were presented via calibrated headphones in a listening room.

Stimuli

Fourteen kettles boiling water were binaurally measured by means of a calibrated head measurement system HMS IV under controlled measurement conditions. Five of the fourteen kettle sounds were chosen and presented twice in a listening experiment.
To provide data about the sound character of the kettles noises, table 1 displays the respective loudness and sharpness values computed by means of the respective German standards. Moreover, the results of Relative Approach analyses are shown. The Relative Approach parameter is related to the perceivable patterns in acoustic signals [9]. This analysis detects specific, obtrusive, attention-attracting noise features. It is often observed that human auditory impression is dominated by patterns, and in such situations largely ignores absolute values [10].

Table 1: Loudness, sharpness and Relative Approach analysis of kettle noise stimuli

Kettle sound	Loudness N_5 according to DIN 45631/A1 (left/right) in sone	Sharpness S_{ave} according to DIN 45692 (left/right) in acum	Relative Approach values (left/right) according to [9]
1	12.5/12.3	1.06/1.04	18.9/18.2
2	16.4/16.7	1.54/1.51	17.4/17.2
3	15.6/15.9	1.24/1.23	19.1/19.1
4	16.2/16.0	1.41/1.41	19.6/19.2
5	22.4/22.1	1.29/1.28	20.5/19.3

2.2.2 Procedure

The sounds were judged by means of a 10-pt unipolar category scale. The scale ranged from „excellent“ to „unbearable“. The relevant five product noise stimuli were played back via headphones only. Later in addition to the playback the test subjects simultaneously looked at a picture, which showed the respective product the sound corresponds to. The subjects were requested to judge only the acoustical stimulus.

2.2.3 Results

Figure 5 shows the sound assessments of the five kettle products under two test conditions. By means of an ANOVA for two-factor repeated measures design with repeated measures on both factors it was observed, as expected, that the main factor product sound was statistically significant.
The product sounds were differently judged with respect to their sound quality and sound character respectively. In contrast, the effect of visual input on the sound assessment was not found to be significant. However, a significant interaction effect was observed ($F_{(4,52)}=4.4$, $p<0.01$**), which means that in dependence of the shown picture the sound quality assessment was differently influenced.
In general, it can be seen that the softest sound with the lowest sharpness was assigned the highest sound quality, whereas the loudest sound achieved the lowest sound quality ratings (see table 1). The other kettle noises exhibit comparable loudness values. The use of Relative Approach analysis results follows the sound quality assessments of the product noises with comparable loudness. This analysis indicates the amount of perceivable patterns in the time and frequency domain and thus might be appropriate to explain the observed judgmental differences in sound quality of the sounds similar in loudness. This parameter seems to be more important than the parameter sharpness, since the second kettle exhibits the

highest sharpness values, but got the second highest sound quality ratings.

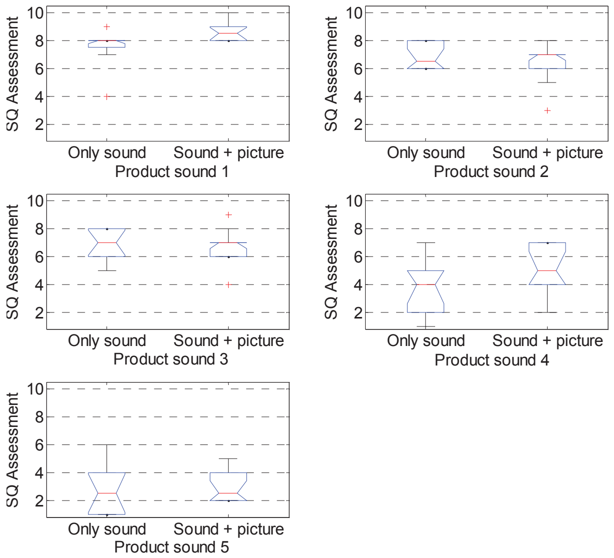

Figure 4. Box-whisker plots: Comparison of sound quality assessments given in laboratory context on a 10-point category scale without and with visual input provided

2.2.4 Discussion

The experimental result suggests that product pictures have an influence on the sound quality judgments of the kettle product sound quality. As expected, the trend and extent of this influence is dependent on the combination of picture and sound, which was indicated by a significant interaction effect. Thus, this effect cannot be understood as a simple offset: sound quality assessments are prone to audio-visual interaction. Although there is a significant correlation between the sound quality ratings of product sound only and the audio-visual stimuli of $r=0.70^{**}$, the assessments are not congruent. It must be pointed out that the test subjects were explicitly requested to judge the acoustical stimulus only. This means that for sound quality optimization audio-visual interaction effects must be taken into account.

3. Conclusions

The presented studies illustrate the influence of contextual aspects on sound quality assessments. Although the existence of context effects is undisputed, the consideration and discussion of such influences in the context of product sound quality are widely neglected.

In general, subjects try to make sense of an unfamiliar experimental environment based on the instructions, cues and feedback they receive [4]. Based on this sense, subjects "develop" their assessments. Thus, any kind of experiment measures more than the experimental objective.

It is difficult to overcome the conflict that uncontrolled variables confound the results in close to reality experiments, but test results are achieved with potentially high ecological validity. For example, larger variations in a close to reality experiments due to uncontrolled confounding variables probably result in a higher variance of data, which causes the risk to accept by mistake the null hypothesis and reject by mistake the research hypothesis. Thus, it is important to acknowledge that test results are often only valid for specific test circumstances and objects investigated, and the detailed exploration of non-acoustical aspects influencing sound quality assessment is mandatory to avoid the identification of pseudo-relationships [11].

References

[1] Blauert, J., Jekosch, U.: Sound-quality evaluation – a multi-layered problem, Acustica and Acta Acustica 83, 5, Hirzel Verlag, 747-753, 1997

[2] House, N., Shively, R.: Placebo method. A comparison of in-situ subjective evaluation methods for vehicles, AES 2000, France, Paris, 2000

[3] Guski, R., Blauert, J.: Psychoacoustics without psychology, NAG/DAGA 2009, Proceedings, Rotterdam, Netherlands, 2009

[4] Marks, L.E., Algom, D.: Psychophysical Scaling, In: Michael H. Birnbaum (eds.). Measurement, Judgment and Decision Making, Academic Press, San Diego, USA, 1998

[5] Haselton, M.G., Nettle, D., Andrews, P.W.: The evolution of cognitive bias, In: D. M. Buss (Ed.). The Handbook of Evolutionary Psychology, John Wiley & Sons Inc., Hoboken, NJ, USA, 2005

[6] Zeitler, A., Fastl, H., Hellbrück, J., Thoma, G., Ellermeier, W., Zeller, P.: Methodological approaches to investigate the effects of meaning, expectations and context in listening experiments, Internoise 2006, Proceedings, USA, Honolulu, 2006

[7] Sköld, A., Västfjäll, D., Kleiner, M.: Perceived sound character and objective properties of powertrain noise in car compartments, Acta Acustica united with Acustica, Vol. 91 (2), 2005

[8] Steffens, J.: Realism and ecological validity of sound quality experiments on household appliances, AIA-DAGA 2013, Proceedings, Merano, Italy, 2010

[9] Genuit, K.: Objective evaluation of acoustic quality based on a relative approach, Internoise 1996, Proceedings, Liverpool, UK, 1996

[10] Genuit, K., Bray, W. R.: Dynamic acoustic measurement techniques considering human perception, ASME International Mechanical Engineering Congress and Exhibition, Chicago, Illinois, USA, 2006

[11] Fiebig, A.: Cognitive stimulus integration of auditory sensations and sound perceptions, Doctoral thesis, TU Berlin, Germany, in press, 2015