Acoust. Sci. & Tech. **29**, 2 (2008)

PAPER

The influence of tone length and S/N-ratio on the perception of tonal content: An application of probabilistic choice models in car acoustics

Hans Hansen* and Reinhard Weber†

C.v.Ossietzky University, Institute of Physics, Acoustics Group, Carl-von-Ossietzky-Str. 9–11, 26111 Oldenburg, Germany

(*Received 30 July 2007, Accepted for publication 18 October 2007*)

Abstract: The rating of tonal content is not just a prevalent issue in the context of sound design but also in annoyance/noise problems. Two physical parameters influencing the evaluated tonal content are tone length and S/N-ratio. In order to integrate these parameters into a reliable, objective measure of the perceived tonal content of interior car noises the following experiments were conducted: Realistic "howling," i.e. the tonal content, in interior car noise were presented with variations in tone length and S/N ratio. Thirteen stimuli were judged by 41 subjects in a complete pair comparison. It was shown that subjective judgments can be modeled by probabilistic choice models quite well. An one-dimensional choice model, i.e. the BTL model, must be rejected in favor of a more general elimination-by-aspect model. The results are the following: (1) The subject's decision aspects are identified by modeling the choice behavior: A non-howling aspect describes the sound evoking a sensation not described by the howling adjective and a pitch-salience aspect indicates the equivalent S/N-ratio of a tonal component. No subjective duration aspect was found. (2) The log-ratio scale obtained yields a constant factor for doubling the length in an interval from 250 ms to 2,000 ms with equal level as well as a constant factor for each increase by 3 dB. Furthermore, the tonal content conception and implications thereof are introduced and discussed.

Keywords: Tonal components, Pitch salience, Subjective duration, Car acoustics, Probabilistic choice models

PACS number: 43.66.Hg, 43.66.Lj, 43.66.Ki [doi:10.1250/ast.29.156]

1. INTRODUCTION

This study is concerned with the investigation of scaling tonal content during a short acoustical setting with interior car noise. Tonal content is understood as the magnitude of a salient pitch extending within a sound sample. Therefore, the physical characteristics tone length and S/N ratio are two obvious moderating factors. The following two aims are pursued in this study: (1) to develop a ratio scale for tonal content and (2) to determine the decision aspects concerning the tonal content. These steps are necessary to validate an objective parameter with the potential to estimate the perceived tonal content of a given sample.

Using the sound quality definition given by Jekosch [1], tonal content can be regarded as a quality feature[1], i.e. it is a part of the perceived nature of an entity and has to be modified in order to change the sound quality.[2] These quality features are subjected to modifying factors [2,3] such as memories, product information, or situational factors, which have been investigated by, for instance, Vaestfjaell [4].

In various contexts the perceived pitches, i.e. the tonal phenomena, are described by numerous onomatopoeic terms. More specifically, in car interior acoustics several different tonal phenomena occur. They are all related to different sources, most notably to the gears, the engine, and the tires and they are described by words such as howling or whining. Normally, these tonal phenomena occur during specific driving conditions, e.g. when the engine is running at a specific RPM, or in the case of the tire-torus-resonance at a specific car velocity. They emerge out of the background noise and after the driving condition is

*e-mail: Hans.Hansen@uni-oldenburg.de

†e-mail: Reinhard.Weber@uni-oldenburg.de

[1] "A recognizable characteristic of an entity that can be denominated." [1]

[2] "Result of an assessment of the perceived auditory nature of a sound with respect to its desired nature." [1]

continuously altered they descent again. Therefore, the tonal phenomena can be considered to be embedded within the remaining car interior noise. The authors decide to evaluate the tonal content within the appropriate context. In this way, the authors emphasize that the stimuli are to be ecologically valid and could be encountered in actual cars.

A tonal content description in terms of a ratio scale is capable of describing the sound character quantitatively. This is one step further describing the most common quality features of sound samples with a quantitative description. An adequate description enables the determination of the perceptual distance between the defined tonal content of a target sound and the actual sound in order to take appropriate countermeasures. This is a paradigm which is commonly used in sound design [5].

The tonal phenomenon under investigation is described as howling. In order to investigate the S/N ratio and tone length dependency or their perceptual equivalent pitch salience and subjective duration dependency on the tonal content a paired comparison experiment was conducted. Actual car noise was filtered to gain realistic stimuli with "howling" content of various magnitude.

The paired comparison data was modeled by probabilistic choice models (Sec. 3.) in order to model the actual choice behavior shown by the test participants [6]. This method was introduced by Zimmer *et al.* [7] into acoustics and was applied to investigate the auditory unpleasantness of different sound sources. Most recently, it has been applied to the evaluation of multichannel reproduced sounds [8]. Using these classes of models a scaling of the underlying evaluation attributes can be achieved. The results of an interview after the experiment enhance these findings. This combination of indirect quantitative and qualitative methods has the advantage of giving a nearly complete picture of the sound perception investigated.

In sum, the study investigates the aspects of the quality feature of tonal content within interior car noise and aims to construct a ratio scale which is based on a sound methodological basis. This is to be achieved by using realistic environmental stimuli, an intuitive, onomatopoeic description of the quality feature, and requiring no assumption about the actual scaling ability shown by test participants.

2. RATING AND SCALING TONAL CONTENTS

The tonal content as a quality feature is affected by various factors. This study expands the tonal content concept beyond a sole pitch salience approach. The pitch salience becomes only an attribute of the quality feature, i.e. it is not satisfactory to calculate a measure of tonal energy and relate it to the amount of noise in a stationary or quasi-stationary noise alone. The contents concept expands to the tone length or subjective duration. Therefore, a tone is considered to be embedded in two dimensions: (1) within the noise across the spectrum and (2) within a noisy context which encompasses the tonal feature before and after its occurrence. The tonal content has more attributes than just pitch salience and subjective duration within a sample. Further attributes would be modulation/fluctuation or the occurrence of several distinct tonal features in different auditory streams at once.

This study investigates the dependency of a specific tonal content on tone length and S/N ratio in a short acoustical scene. In more detail, a howling component within a car interior noise will be evaluated. There are different solutions within different fields which cover the scaling problem of the subjective evaluation of pitch salience, e.g. [9–12]. These parameters do not assess the tonal content of an acoustic scene in total, but they describe the pitch salience in a stationary or quasi-stationary noise.

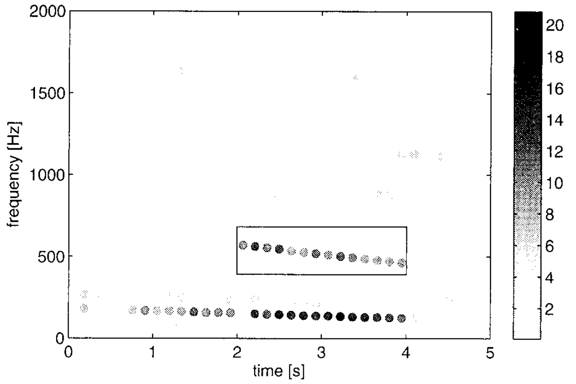
Figures 1(a) and 1(b) show an analysis of tonal content using DIN 45681 [10], a norm to detect tonal components in noise and to determine a 'tone adjustment' for the assessment of noise immissions. The norm uses basically the level above masked threshold within the neighboring critical band (CB). Originally, however, it excluded every other tonal energy within the CB. Therefore, DIN 45681 was modified and the masked threshold was corrected by calculating the excitation of the tonal components [13].

Regarding low frequency components within car interior noise, Hansen *et al.* [ibid] stated that a booming sensation is not connected to a distinct tonal feature. The authors have shown that onomatopoeic description for a tonal feature reflecting a specific source, e.g. howling or whining can be related to the overall description tonal content. Therefore, in the present experiment the description howling is used though in order to increase ecological validity of the evaluation procedure.

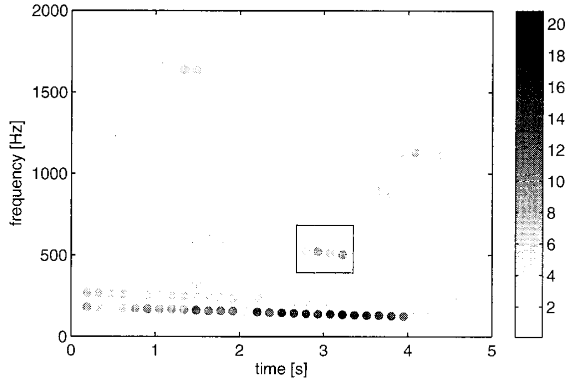
There seems to be at least two major physical dimensions affecting tonal content: duration and S/N ratio. The S/N ratio can be identified with the saliency problem while the duration refers to the duration of the salient signal within the given noisy context. In front of the background, the scaling of a single sinusoid is no longer a valid question as there is not a background/context which this stimulus contents. This experiment aims to analyze the trade-off relationship between duration and S/N ratio the decision aspects which underly the judgment about this specific tonal content.

3. PAIRED COMPARISON METHOD AND PROBABILISTIC CHOICE MODELS FOR SCALE CONSTRUCTION

In order to generate a ratio scale of noises according to their tonal content, the paired comparison as an indirect



(a) Stimulus S_{10}; tone length: 2 s, S/N ratio: no reduction



(b) Stimulus S_5; tone length: 0.5 s, S/N ratio: -3 dB

Fig. 1 Analysis according to the modified DIN 45681 [13]. The diagrams show the extracted tonal components which exceed masked threshold at their appropriate critical bandwidth. As a measure of excess the ratio of tonal energy and the noise energy corrected by masked threshold [in dB] are given. The tonal energy at the specific frequency at the analyzed point in time of the extracted tonal components is characterized via differently shaded dots. Note, frequencies below 200 Hz are not considered as tonal components as such and are more related to booming sounds [ibid].

scaling method is used [14]. It is indirect in the sense that the test participant does not apply any rating value to the stimuli. Afterwards, the scale values are calculated from the test participant's response. Compared to direct scaling methods such as categorical scaling or magnitude estimation, the indirect approach has two major advantages: (1) The participant can focus on the stimuli's perceptual properties, because it is not directly involved in the actual scaling. (2) At test participant's capability of scaling is not required [15]. Furthermore, the actual decision strategies can be modeled by using probabilistic choice models.

The following steps are necessary in order to finally generate a ratio scale, a paradigm used by Zimmer *et al.* [7] or recently by Choisel and Wickelmaier [8]:

Pair Comparison Test Presentation of two stimuli in a random order, the test participant chooses one stimulus with respect to a specific criterion.

Cumulative Pair Comparison Matrix Generating a cumulative pair comparison matrix, i.e. summing up all preference matrices of all test participants.

Decision Model Modeling the decision by assigning common aspects to the stimuli (Sect. 3.1.).

Model Choice Testing the respective model by not only using a goodness-of-fit test but also by checking the respective prerequisites (Sect. 3.2.).

3.1. Decision Models

In the following subsection two decision models will be briefly compared regarding their capability of coping with perceptual aspects: the Bradley-Terry-Luce (BTL) [16] and the Elimination-By-Aspect (EBA) [6] model.[3]

3.1.1. The BTL model

The choice axiom underlying the BTL model is given by:

$$p_{xy,\mathrm{BTL}} = \frac{v(x)}{v(x) + v(y)} \tag{1}$$

$p_{xy,\mathrm{BTL}}$ is the preference probability, i.e. the probability of choosing stimulus x over y, while $v(x)$ and $v(y)$ are the stimulus' scale value of the attribute in question. The scale values are obtained out of the preference probability matrix, i.e. the results in Table 2 divided by test participant number. Though a simple approach in the sense that every stimulus is modeled by a single parameter it has a major drawback. There has to be context independence, i.e. the perceptual aspect has to be the same in any single trial of the paired comparison test. The context independence is at stake if the perceptual aspect under investigation is highly similar within a given pair, because other discriminatory aspects are used for comparison as the actual aspects turns out to be almost equal. Please refer to Sect. 3.2.2. for a formal prerequisite's formulation.

3.1.2. The EBA model

One possibility to overcome the difficulties of multidimensional decision criteria and context dependencies is to choose a more general decision model as the Elimination-By-Aspects (EBA) model [6]. This model assumes the "elimination" of equal aspects of stimuli as a decision basis, i.e. if two stimuli have the same attribute (aspect) this common attribute is eliminated in the preference decision. The preference probability[4] $p_{xy,\mathrm{EBA}}$ for choosing x over y is given by:

[3]Notation through this section as in [17]

[4]Only binary choice probabilities are analyzed.

$$p_{xy,\mathrm{EBA}} = \frac{\sum_{\alpha \in x'|y'} u(\alpha)}{\sum_{\alpha \in x'|y'} u(\alpha) + \sum_{\beta \in y'|x'} u(\beta)} \quad (2)$$

where α and β are the aspects of the stimuli x and y. The set of aspects belonging to x and y is denoted by x' ($x' = \{\alpha : \alpha \text{ aspect of } x\}$) and y' ($y' = \{\alpha : \alpha \text{ aspect of } y\}$). The set $x'|y'$ contains all aspects of x that are no aspects of y and the contrary holds for $y'|x'$. The function $u(i)$ assigns a scale value for each stimulus' aspect. In the general EBA case the sum is taken over the set $x'|y'$ of aspects. Here is $x'|y'$ the set which includes all aspects describing stimulus x but not y. The values $u(\alpha)$ are the model's parameters. The model parameters $u(\alpha)$ are calculated by solving the Eq. (2) numerically. $p_{xy,\mathrm{EBA}}$ are determined by the preference matrix. In such a way, it is possible to investigate the decision strategies used by the test participants, even if more than one attribute is driving the decision process.

In the case of the BTL model, there is only one aspect per stimulus and in this case the equation 2 reduces to equation 1. Hence, the BTL model is a special case of the EBA model in which the decision is only based on one attribute/aspect of every stimulus.

3.2. Model Choice and Stochastic Transitivities

An advantage in using probabilistic choice models is that models which have been generated *a priori* cannot only be tested by a goodness-of-fit test, but also have defined prerequisites which have to be met.

3.2.1. Goodness-of-Fit test

The goodness-of-fit test is realized by testing the log-likelihood ratio. This ratio, which is approximately χ^2 distributed, is based upon the ratio of the model likelihood L and saturated solution likelihood L_{sat}:

$$\chi^2 = -2 \log\left(\frac{L}{L_{\mathrm{sat}}}\right) \quad (3)$$

The likelihood function L is determined by the fact that the paired comparison matrix can be seen as $\binom{n}{2}$ binomially distributed random variables is given by:

$$L = \prod_{i>j} \pi_{ij}^{N_{ij}} (1 - \pi_{ij})^{N_{ij}} \quad (4)$$

where i and j are the row and column indices of the paired comparison matrix (Sect. 5.1); N_{ij} being its elements. The π_{ij}, the estimates of the preference probability, are obtained by solving equation 1 or 2 numerically. The likelihood function L_{sat} of the saturated model is estimated by setting $\pi_{ij} = \frac{N_{ij}}{N_{ij}+N_{ji}}$.

Usually a minimum solution for L cannot be obtained analytically as the configuration of $u(\alpha), u(\beta), \ldots$ that minimizes L has to be found. In this study this is numerically done with a *Matlab* script by Wickelmaier and Schmid [17].

3.2.2. Different stochastic transitivities as models' prerequisites

As the prerequisites of both models, BTL and EBA, are known they can be checked to suggest modeling approaches. An important prerequisite is the stochastic transitivity. The *weak stochastic transitivity*,

$$((p_{xy} \geq 0.5) \wedge (p_{yz} \geq 0.5)) \Rightarrow p_{xz} \geq 0.5 \quad (5)$$

p_{xy} being the probability of choosing x over y, is the condition for ordering the stimuli in a one dimensional, ordinal way with respect to, for instance, 'howling,' as in the present paper. Thus the preference probability matrix should not violate this principle.

The *medium stochastic transitivity*,

$$((p_{xy} \geq 0.5) \wedge (p_{yz} \geq 0.5)) \Rightarrow p_{xz} \geq \min\ p_{xy}, p_{yz} \quad (6)$$

is a prerequisite for the EBA modeling approach. The BTL model implies an even more restrictive regime: the *strong stochastic transitivity*,

$$((p_{xy} \geq 0.5) \wedge (p_{yz} \geq 0.5)) \Rightarrow p_{xz} \geq \max\ p_{xy}, p_{yz} \quad (7)$$

and it is therefore necessary to check the preference matrix whether a BTL approach is viable [18].

Thus, it is possible to evaluate the models not only by a statistical test, but also by checking the prerequisites which should be met.

4. EXPERIMENTAL DESIGN

4.1. Stimuli

In order to maintain ecological validity, a realistic interior car noise in a "coasting" condition was chosen as the basic sound. A howling component (19th engine order, about 570 Hz, see Figure 1) is clearly audible. This sound was binaurally recorded using an artificial head at the front right position.[5]

All stimuli versions were obtained through digital filtering of the original recorded sound sample, i.e. the howling component was either removed completely, or removed and afterwards the component was blended in again but modified in its tone length and S/N ratio. In this way, the background noise was held constant with a level of 54.0/57.0 dB(A) (left/right). The overall level did not change significantly by adding the weak tonal component. The maximum for stimulus 10 is 54.2/57.1 dB(A) (left/right). The four different tone lengths are: 250 ms, 500 ms, 1 s, and 2 s. The howling component was presented at the original level and at reduced levels (−3 dB, −6 dB). Combining these parameters yields 12 stimuli (see

[5]Recorded by *Akustikstudio* of the *Mercedes Benz Technology Center* in Sindelfingen, Germany.

Table 1 Experimental Stimuli. The different tone lengths and S/N ratios of the tonal content (howling) are shown for each stimulus used in the experiment. The stimulus 13 (indicated by a ⋆) does not contain any howling.

Stimulus S_i	Tone length [ms]	S/N ratio modifications [dB]
1	250	−0
2	250	−3
3	250	−6
4	500	−0
5	500	−3
6	500	−6
7	1,000	−0
8	1,000	−3
9	1,000	−6
10	2,000	−0
11	2,000	−3
12	2,000	−6
13⋆	na	$-\infty$

Table 1). As the 13th stimulus a condition without any howling component is included.

Figure 1 shows human-related analyses of the prominent tonal content using a modified German Industry Norm DIN 45681 for extracting tonal components [10]. By setting the tonal contents in relation to its background the graphical representation intends to fit the actual perception in an appropriate way.

4.2. Test Procedure

For a complete paired comparison test the participants have to be instructed to choose one stimulus over the other with regard to the specific criterion, i.e. howling. The question asked is "Which sound is more howling?". In order to be close to the participant's own onomatopoeic description howling is chosen instead of a more abstract question concerning tonal content. In a semantic differential study, Hansen *et al.* [13] have shown that howling can be regarded as a more specific form of the adjective pair tonal/not tonal. More precisely, the two adjective scales in this study referring to specific tonal content, i.e. howling and whining, span a plane in the perceptual space where the adjective scale tonal/not tonal is the bisecting line between both.

The entire experiment has three distinct phases: a training and orientation phase, a main test phase, followed by a concluding interview.

Orientation Phase At the start the investigator instructed the participant to read the instruction presented on a screen. The participant listened to five stimuli (S_9, S_{13}, S_5, S_1, S_{12}) for orientation. As the experiment is designed as an analytical listening experiment, the investigator made sure that the stimuli were perceived in a way that the howling was correctly identified.

Training Phase The training experiment started with stimuli: S_1, S_5, S_{10}, and S_{13}. The participant then had to press buttons on a screen in order to start the stimuli presentation. The participants were allowed to choose the listening order of the stimulus pair and could repeat listening to the stimuli as many times as they needed to. After this training phase the investigator answered further questions with regard to the stimuli apart from a direct description of the modifications.

Main Test Phase This phase was identical to the training phase, but all stimuli were used thus resulting in 78 paired comparisons.

Concluding Interview The participants were interviewed in order to gain an overview about the decision aspects and first ideas about a modeling of the decision aspects. In this way, the participant may freely describe any thought crossing their minds during the experiment. If the participant did not directly describe some strategy, the investigator will lead the interview to a strategy description.

The whole experiment lasted about 45 min per participant.

4.3. Experimental Set-Up

The experiment was performed in a sound-proof room at Oldenburg University. The stimuli were presented via headphones (Sennheiser HE 60 & Preamplifier Sennheiser HE 70). The DAT recorder Sony 57ES was used as the D/A converter connected to a PC. The PC was also used to run a *matlab* paired comparison developed in the acoustic group.

4.4. Participants

Four female and 38 male participants participate in this experiment. The mean age was (30.9 ± 8.2) years. All participants did not have any reported history of hearing loss.

5. RESULTS

The resulting cumulative paired comparison matrix is shown in Sect. 5.1., cf. Table 2, and the outcome of the interviews are summarized in Sect. 5.2. Reliability tests concerning inter- and intra-subject reliability are performed to exclude unreliably answering participants and to check the experimental results' internal consistency. Thereafter, the stochastic intransitivity and the probabilistic choice model approach are evaluated. After choosing an appropriate decision model the ratio scale is constructed.

5.1. The Cumulative Paired Comparison Matrix

The individual responses were pooled in a cumulative paired comparison matrix (Table 2). This matrix is the basis for further analysis. The cell entries are the number of

Table 2 Cumulative paired comparison matrix. The cell entries are the number of participants who judged a stimulus in a column less howling than a stimulus in the corresponding row, e.g. stimulus 11 is judged less howling than stimulus 7 by 9 participants. The maximum number is 41, the number of reliable participants.

Stimulus S_i	1	2	3	4	5	6	7	8	9	10	11	12	13
1	—	37	41	2	18	37	1	8	27	0	1	12	41
2	4	—	40	1	2	32	0	2	15	0	1	4	41
3	0	1	—	2	1	7	0	0	5	0	0	1	37
4	39	40	39	—	38	37	1	16	37	0	4	17	40
5	23	39	40	3	—	38	1	4	23	0	0	8	41
6	4	9	34	4	3	—	0	2	9	2	2	0	38
7	40	41	41	40	40	41	—	36	40	1	9	31	41
8	33	39	41	25	37	39	5	—	34	0	4	17	40
9	14	26	36	4	18	32	1	7	—	0	0	3	37
10	41	41	41	41	41	39	40	41	41	—	35	39	40
11	40	40	41	37	41	39	32	37	41	6	—	35	41
12	29	37	40	24	33	41	10	24	38	2	6	—	39
13	0	0	4	1	0	3	0	1	4	1	0	2	—

participants who judged a stimulus in a column less howling than stimulus in the corresponding row. If divided by the maximum number 41, the number of reliable participants, the probabilities of choosing x over y are obtained. Also a number of vacancies occur, i.e. frequencies in Table 2 being 0 or 41. (One test participant had to be excluded due to low reliability (Sect. 5.3).)

5.2. Interview Summary

The paired comparison test was followed by a short interview which was conducted in an open manner (Sect. 4.2.). The results can be summarized as follows:

Duration 25/41 participants mentioned duration as the first decision criterion especially when confronted with a trade-off scenario between duration and S/N ratio.

S/N ratio Only 5/41 participants are mentioned the saliency of the tonal content as their main criterion.

Recognizing Variations 41/41 participants reported the two dimensions in which the stimuli vary.

Equal Pair 10/41 participants insist that there are some pairs which are perceived as totally equal.

No rational strategy 5/41 participants reported a strategy but judging according to their "gut feeling."

Short howling A few participants did not consider the short howling as real howling, i.e. they did not consider the term howling appropriate to describe the differences of some of the stimuli.

5.3. Inter- and Intra-subject Reliability

A common method in evaluating paired comparison data with regard to intra-subject reliability is to relate the number of circular triads participants obtained in their process of judgment to the number of possible circular triads [14]. A circular triad is defined as the following: Let x, y, z be binary decisions with regards to three objectives. If the participants judge in the following manner $(x > y) \wedge (y > z) \wedge (z < x)$. It can be regarded as inconsistent. The ratio ξ between actual and possible circular triads is approximately χ^2 distributed. In this actual test participants have to be rejected, if they have more than 59 actual circular triads ($p = 0.05$). As a result only one participant is excluded from further investigation. The median of ξ is 0.022, i.e. 2 circular triads.

Kendall's accordance coefficient constitutes the level of inter-subject reliability. It measures whether the concordant judgments are within the scope of chance (for a detailed derivation please refer to [19]). The accordance coefficient of the matrix found in Table 2 is 0.71 [$\chi^2(84) = 2409$, $p < 0.001$], which means that the concordant judgments are not a result of chance.

The highly concordant judgment is in line with the post-experiment interview, in which most of the persons expressed the ease of most decisions. As the circular triads are very low in most cases, the person tended to judge in a transitive way. Both the good inter- and intra-subject reliability and the data from the post-experiment interview are a good foundation for further analysis.

5.4. Stochastic Transitivities

In Sect. 3.2. the condition of stochastic transitivity was introduced. Applying this to the preference probability matrix, i.e. the cumulative paired comparison matrix divided by the participant number, yields the following results:

- The *weak stochastic transitivity* is never violated. Therefore, all stimuli can be ordered in a one-dimensional way (cf. Eq. (5)).
- The *medium stochastic transitivity* is violated 6 out of 286 possible times. This suggests a possible modeling using the EBA approach.
- The *strong stochastic transitivity* is violated 76 out of 286 possible times casting doubt on possible BTL solutions as this transitivity is a BTL model's prerequisite.

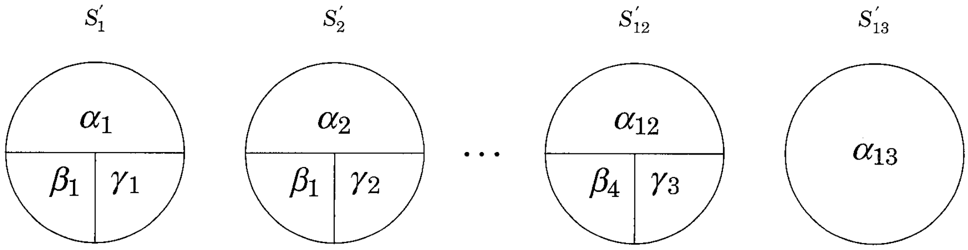


Fig. 2 Elimination-by-Aspect model (I). The figure illustrates the sets of aspects S_i' of the stimuli S_i, $i = 1, \ldots, 13$. As a first approach, the decision aspects are modeled along the varying physical parameters. At first, an aspect is introduced for every stimulus $\alpha_{1,\ldots,13}$. Secondly, aspects are introduced which have an equal tone length the decision aspect subjective duration $\beta_{1,\ldots,4}$ and an equal S/N ratio the decision aspect pitch salience $\gamma_{1,\ldots,3}$ (Refer to Table 1 for the common physical parameters). Stimulus S_{13} does not have additional aspects as it has no howling. The EBA model can be explained by looking at the comparison between stimulus S_1 and S_2. The model states that S_1 and S_2 have a common aspect β_1. This aspect is "eliminated" and not used for the decision.

5.5. Modeling Paired Comparison Data by BTL and EBA Models

Now different decision models are fitted to the content of the paired comparison matrix. The parameter estimation is done numerically as outlined in Sect. 3.1.

5.5.1. BTL model

The most simple approach is to assume one decision attribute α_i for each stimulus, i.e. the BTL modeling approach. Yet a goodness-of-fit test shows (Sec. 3.2.1.), that the fit is not significant [$\chi^2(66) = 155.7$, $p < 0.001$].[6] This corresponds to the violation of the strong stochastic transitivity reported in Sect. 5.4.

5.5.2. EBA model (I)

Apart from the BTL approach several models can be derived a priori by knowing the stimuli's physical parameter/attribute variations and by analyzing the interview. Assigning similar perceptual aspects such as the same subjective duration or the same pitch salience to corresponding equal physical parameters leads to several models tested as these aspects were mentioned by all participants in the interview.

In order to give a reasonable and sufficiently general example, one could model every stimulus by a single parameter $\alpha_{1,\ldots,13}$ and then add additional parameters for equal subjective duration $\beta_{1,\ldots,4}$ and pitch salience $\gamma_{1,\ldots,3}$. In Fig. 2 a sketch of the decision aspects is drawn and the aspect assignment is documented in Table 3.

Table 3 EBA I and EBA IIa model's decision aspects. The model EBA I is a model closely connected to the stimuli's physical variation (cf. Table 1). The EBA II models introduces an aspect δ which is connected to a non-howling perception. The model EBA IIa includes the aspect subjective duration β_i.

Stimulus S_i	decision aspects EBA I			decision aspects EBA IIa		
1	α_1	β_1	γ_1	α_1		δ
2	α_2	β_1	γ_2	α_2		δ
3	α_3	β_1	γ_3	α_3		δ
4	α_4	β_2	γ_1	α_4	β_2	
5	α_5	β_2	γ_2	α_5	β_2	δ
6	α_6	β_2	γ_3	α_6	β_2	δ
7	α_7	β_3	γ_1	α_7	β_3	
8	α_8	β_3	γ_2	α_8	β_3	
9	α_9	β_3	γ_3	α_9	β_3	
10	α_{10}	β_4	γ_1	α_{10}	β_4	
11	α_{11}	β_4	γ_2	α_{11}	β_4	
12	α_{12}	β_4	γ_3	α_{12}	β_4	
13	α_{13}			α_{13}		δ

5.5.3. EBA model (IIa + b)

The EBA model (I), though a first reasonable EBA approach, did not suffice the goodness-of-fit criterion. Only after stimuli $S_{1\text{–}3,5,6,13}$ are modeled with a common aspect δ, the fit gets significantly better. This decision aspect is hinted by the interview results in which a couple of participants report that some stimuli are not considered as howling (Sect. 5.2.). The models which include these aspects are considered as the EBA (II) model throughout this paper.

Now, two specific models are directly compared:

EBA IIa-subjective duration In this model only the common decision aspect subjective duration $\beta_{2\text{–}4}$ is introduced to the stimuli including the 3 longest tone lengths, i.e. stimulus number larger than 3. Table 3 introduces the models' decision aspects. The common decision aspect $\beta_{2\text{–}4}$ improves the fit even more [$\chi^2(62) = 105.2$, $p < 0.001$].

EBA IIb-pitch salience A different approach, which turned out to be the best, is to model only the decision aspect pitch salience $\gamma_{1\text{–}3}$. Table 4 shows the structure of the decision aspects within this model. The fit is significant on a $p = 0.01$ level [$\chi^2(62) = 90.9$, $p = 0.01$]. In this way, the EBA IIb model is chosen in order to generate a ratio scale.

[6] The null hypothesis is that the model under scrutiny holds, in this case the BTL model.

Table 4 The final EBA IIb model's decision aspects & normalized log-scale-values for each stimulus. As in the EBA IIa model the non-howling aspect δ of perception is modeled. The model includes the pitch salience aspect γ_i as well. As a result of this final, significant EBA model log-ratio scale values are calculated for each stimulus (cf. Fig. 3).

Stimulus S_i	Decision aspects EBA IIb			Log-scale value howling
1	α_1		δ	0.86
2	α_2		δ	0.31
3	α_3		δ	0.03
4	α_4	γ_1		1.57
5	α_5	γ_2	δ	0.87
6	α_6	γ_3	δ	0.13
7	α_7	γ_1		2.14
8	α_8	γ_2		1.51
9	α_9	γ_3		0.69
10	α_{10}	γ_1		3.19
11	α_{11}	γ_2		2.49
12	α_{12}	γ_3		1.60
13	α_{13}		δ	0.00

5.6. Ratio Scale Construction

The model EBA IIb including the aspects $\gamma_{1\text{–}3}$ modeling the S/N ratio is chosen according to Sec. 5.5. as this model yields a satisfactory fit. Since the solution is a ratio-scale, one scale value can be chosen arbitrarily. The scale value of stimuli S_{13} is set to 1. Furthermore, the log_{10}-function is applied to the scale values resulting in a value of 0 in the case of stimuli S_{13}. Only differences are meaningful on this log-transformed scale. Figure 3 shows a plot of the logarithmic scale values including the calculated standard errors for easy comparison.

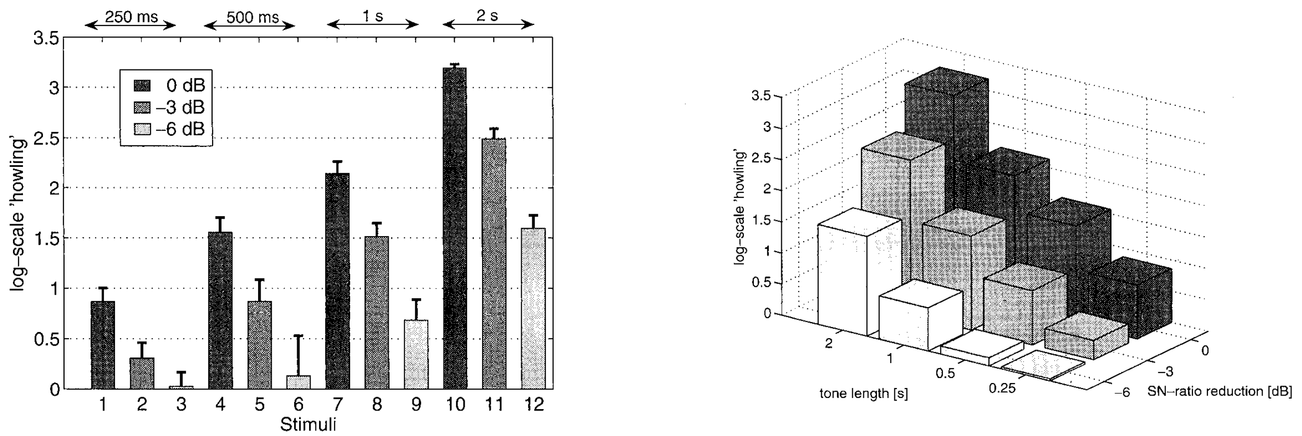
6. DISCUSSION

After an appropriate decision model has been found the implications of the decision-model aspects are discussed in the light of general scaling issues for tonal content. Furthermore, the howling scale which actually resulted is considered.

6.1. Decision Aspects

In order to determine the relevant decision aspects in a valid model several approaches are tested. Concerning the experiment conducted, the BTL approach, a common paradigm, does not lead to a valid decision aspect modeling. At first, the decision aspects were modeled only in concordance to the physical parameters, i.e. tone length and S/N ratio. This approach had to be abandoned in favor of models which take the test participants' description of the stimuli into consideration.

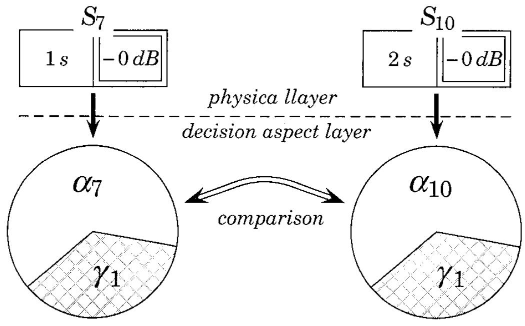
In this way, a satisfactory fit could be reached with the EBA IIb model approach ($p = 0.01$) by introducing a non-howling aspect δ and a pitch salience aspect $\gamma_{1\text{–}3}$ as additional decision aspects (cf. Table 4). The first aspect is labeled as a result of the interviews as non howling. The stimuli which are described by the additional aspect non-howling include an identifiable pitch, but the onomatopoeic description of howling does not fit here with these rather short subjective durations in the 5 s noisy context. Nevertheless, the transition to fully howling might be continuous as stimulus S_9 has a smaller scale value than some of the stimuli within the non-howling group, but it is considered rather faintly howling. The non-howling aspect stresses the



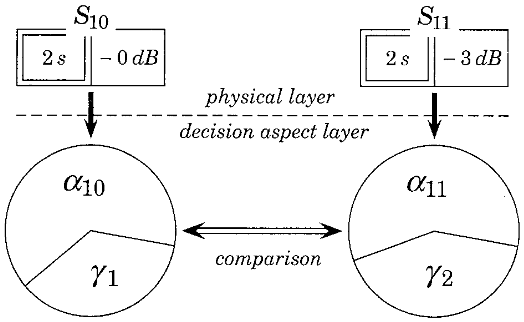
(a) The figure shows the log-scale values of Table 4 along with the standard error of mean. A similar decrease on the howling scale can be observed by reducing the S/N ratio within each tone length group.

(b) Here the log-scale values are plotted along the 4x3 experiment design setting. The 3D-bar plot shows a trade-off between 3 *dB* increase of S/N ratio and doubling the tone-length. The independent physical parameters do not necessarily translate into independent perceptual parameters (Sect. 6.1).

Fig. 3 Log-scale values of howling. The scale values are generated by fitting a probabilistic choice model EBA IIb (cf. Table 4) to the paired comparison matrix (cf. Table 2) in order to generate a ratio scale. The decision aspects are listed in Table 4. Stimulus S_{13} is arbitrarily set to 0.



(a) Comparison based on extractable dimensions - the crosshatched aspect γ_1 is "eliminated" from the decision process.



(b) Comparison based on overall similarity - the equal *physical* condition is not represented by a decision aspect. Therefore, there is no aspect "elimination".

Fig. 4 Symbolic sketch of decision aspects. In order to illustrate the decision aspects which are obtained by using probabilistic choice models two exemplary decisions are sketched. The physical parameters are drawn in the boxes (physical layer), while the circles hold the resulting decision aspects (decision aspect layer) as introduced in Table 4.

difference between pitch salience and scaling of tonal content in a given ecological context, in this case a driving situation. Although the pitch salience is known to reach its maximum at about 200 ms the scaling of the tonal content does not saturate [20]. Even more, in this environmental situation the stimuli having a rather short duration, i.e. 250 ms or 500 ms, evoke more a sensation of a "peep" than of howling or a distinct tone.

The second aspect is the aspect of similar S/N ratios. In order to illuminate the aspect's meaning in the decision process, two cases are discussed (Please refer to Fig. 4 for sketches):

- Figure 4(a) shows the exemplary comparison between stimuli S_7 and S_{10}. These stimuli feature the same S/N ratio, −0 dB, but have different tone lengths, 1 s and 2 s. In this case the common S/N ratio aspect is "eliminated" from the decision. The test participant chooses the stimulus only according to its remaining differences, in this case the subjective duration. Therefore, the decision is based on an extractable dimension.
- In a second exemplary case (cf. Fig. 4(b)), two stimuli featuring the same tone length, 2 s, but different S/N ratios, −0 dB and −3 dB are compared. In this case the similar subjective duration is not "eliminated" as it is not regarded as a distinct aspect by the test participants. The decision is done in comparison to the overall similarity.

This leads to the conclusion that either solely the subjective duration or the combination of subjective duration and pitch salience is used in this judgment. Turk and Sawusch [21] predicted a dependency of solely length or length and level in prominence cues within speech perception. They state an *asymmetric integrality* of the two varied dimensions in the sense of Garner [22].

Following the argumentation in Figure 4, these could be the case for the judgment of tonal content as well. Nevertheless, according to Ashby and Townsend [23], a difference between perceptual and decisional independence has to be made. The modeling of the decision aspects does not allow differentiating these kinds of independences. Although out of logical reasons no such inferences can be made, the integrality stated would have to be allowed for indices describing tonal phenomena.

6.2. Howling Ratio Scale Construction

Analyzing the log-ratio scale values in Fig. 3 in more detail. The reduction of the S/N ratio by 3 dB leads to a constant reduction of the howling value of about 0.6 (factor of 4) across all tone length. Or in other words, the difference between the highest S/N ratio and the lowest S/N ratio of a specific tone length seems to be constant. An exception is the shortest tone length.

A trade-off between 3 dB increase of S/N ratio and doubling the tone length becomes apparent. The stimuli $S_{1,5,9}$, $S_{4,8,12}$, and $S_{7,11}$ have similar ratings on the ratio scale. The same holds for the differences between a tone length of 0.5 s and a tone length of 2 s and specific S/N ratios. Figure 3(b) displays this behavior in a "step-like" structure accordingly. The constant trade-off suggests a linear correlation between the tonal energy and the log-ratio scale of the tonal content. This leads to a preliminary conclusion that the tonal energy can formally predict the judgement of the tonal content in this experiment, as depicted in Fig. 4(b). The contents in Fig. 4(a) shows something different. If the pitch salience is perceived as equal, it will be "eliminated" from the decision. Regarding this decision, the tonal energy is not decisive as solely the tone length is used. The non-howling aspect δ emphasizes the limit of a tonal energy approach even further as it induces a quality change, which is not covered by this sole quantitative measure.

Nevertheless, there are some inherent limitations which are not in the scope of this experiment. The specific loudness of a tone within a noise masker shows a linear relationship between log-loudness and tone level in noise for high S/N ratios (cf. Moore *et al.* [24]). In the present experiment the S/N ratio is on the one hand rather low, but on the other hand only varied by 6 dB in total. In this way, the constant increase between S/N ratio and the log-ratio scale is assumed to be a linear approximation of this particular S/N ratio level.

Furthermore, the experiment cannot determine whether the absolute or relative duration of the tonal component is used in order to determine the tonal energy as they covariate. This question will be investigated in further experiments.

The span between stimulus S_{13} and S_{10}, the extreme values, is about 3, i.e. a factor 1,000. Such high factors are typically observed above masked threshold where the loudness of a partially masked tone in noise increases rapidly with the level of the tone [ibid.]. This drastically illustrated the enormous perceptual distances of tonal perception.

The howling perception increases within the range of 0.25 s and 2 s. It has to be stated that this refers not to the pitch salience at a specific point in time, but rather to the tonal content perception of the stimuli as a whole. The maximum pitch salience is reached after approximately 200 ms [20], i.e. in this experiment concerning the tone length the maximum of relative pitch salience is reached and it is therefore only varied by different S/N ratios. In this way, the 200 ms is a lower limit for the constant trade-off, between the tone length and the S/N ratio as the former will additionally influence the latter. As shown in Sect. 6.1, the non-howling aspect δ even sets the limit for the howling perception in the region of 250 ms–500 ms as the sound does not fit the onomatopoeic description anymore.

Kuwano and Namba [25] and Namba *et al.* [26] investigate the relationship between overall and continuous judgment of sounds. This is related to the question how the instantaneous pitch salience is related to the overall judgment of tonal content. More specifically, Kuwano and Namba [25] found in the case of loudness a 2.5 s time constant, which relates to the psychological present. In this way, all tone lengths used in this present experiment can be regarded as "within the present," i.e. the pitch salience within the total tone length should contribute to the overall judgment. Hence, 2.5 s is expected to be an upper time limit for the constant trade-off. Moreover, it is yet to be investigated how different pitch salience levels of a tonal percept contribute to this overall judgment.

It can be concluded that the perception of tonal content, i.e. referred to as howling in this context, has a huge dynamic range and might include several phenomena at once: the non-howling components, the distinct components clearly distinguishable from background noise, and tonal phenomena which seem to be part of the noise, but are rather weak (e.g. stimulus 9). Another limit can be set at the border between (1) a noise containing a tone and (2) a tone with added background noise. From the conception explicated in Sect. 2 the definition of tonal content would render useless in the latter case (2) as there is nothing to be contained. The limits of these perceptual classes within the perception and evaluation of tones in noise can be a basis for a sustainable index of tonal content.

7. CONCLUSION

The following conclusion can be stated:

- A definition of tonal content is given, expanding the concept of a stationary pitch salience toward the embedding within two dimensions: (1) the noise across the spectrum and (2) a noisy context which encompasses the tonal feature before and after its occurrence within a given sound sample.
- The test participants' choice behavior can be modeled, when the more general EBA modeling approach is used. Apart from a distinct subjective duration aspect a non-howling aspect is discovered, which is also reported by the test participants in a post-experimental interview. This result might hint at an *asymmetric integrality* of subjective duration and pitch salience as found by Turk and Sawusch in speech perception along these dimensions [21]. The exact level, i.e. perceptual or decisional, at which this "integrality" is located is yet to be determined.
- As the tone length doubles within the range between 0.25–2.00 s, the perception of tonal content is increased by a constant factor. The longer the tone duration, the higher the tonal content is judged. This constant factor is limited by two boundaries: A tone length shorter than 200 ms has an impact on pitch salience [20], i.e. interaction between the tone length and the S/N ratio is detected in this low duration region. Although the howling perception, accessed by a specific onomatopoeic description, might be evoked between 250–500 ms, a non-howling aspect is discovered. The duration upper limit for the constant factor could be 2.5 s, the psychological present, which according to [25] plays a major role in the overall loudness judgment of sounds fluctuating in level. The specific loudness of a tone within a noise masker characterized by a high S/N ratio can be described by a linear relationship between log-loudness and tone level (cf. Moore *et al.* [24]). For noise characterized by low S/N ratio levels, the relationship between log-loudness and tone level is even highly non-linear. The constant factor obtained by decreasing tonal contents via an S/N ratio reduction is therefore assumed to be a linear approximation of the respective function as the level is only varied by 6 dB.
- By reducing either the tone length or the S/N ratio, the

howling perception, i.e. the tonal content, is decreased by a constant factor. The notion of a specific relation to tonal energy is a formal one. In comparing two stimuli of equal S/N ratios, an intuitive counter-example was demonstrated. As the pitch salience is "eliminated" from the decision process, a simple comparison between the different subjective durations provides a new decision basis. Also, the non-howling aspect modeled for low tonal content irradiates the limitation of a exclusive tone energy approach for tone lengths between 250–500 ms.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Dr.-Ing. U. Letens (DaimlerChrysler AG) who provided the original car sounds and shared his knowledge and experience in the field of car acoustics.
We also acknowledge the support of S. Wolter who did a tremendous effort in finding test participants and conducting experiments.

The authors further recognize the efforts of P. Rong in developing a paired comparison matlab-script.
Furthermore, we would like to thank F. Wickelmaier for the fruitful personal communication.

Finally, we would like to thank Cara H. Kahl (University of Hamburg) for her editorial suggestions.

REFERENCES

[1] U. Jekosch, "Basic concepts and terms of 'quality,' reconsidered in the context of product-sound quality," *Acta Acust. Acust.*, **90**, 999–1006 (2004).
[2] J. Blauert and U. Jekosch, "Concepts behind sound quality: Some basic considerations," *Proc Internoise 03* (2003).
[3] J. Blauert and U. Jekosch, "Sound-quality evaluation—A multi-layered," *Acta Acust. Acust.*, **83**, 747–753 (1997).
[4] D. Vaestfjaell, "Contextual influences on sound quality evaluation," *Acta Acust. Acust.*, **90**, 1029–1036 (2004).
[5] W. Keiper, "Sound quality evaluation in the product cycle," *Acta Acust. Acust.*, **83**, 784–788 (1997).
[6] A. Tversky, "Elimination by aspects: A theory of choice," *Psychol. Rev.*, **79**, 281–299 (1972).
[7] K. Zimmer, W. Ellermeier and C. Schmid, "Using probabilistic choice models to investigate auditory unpleasantness," *Acta Acust. Acust.*, **90**, 1019–1028 (2004).
[8] S. Choisel and F. Wickelmaier, "Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference," *J. Acoust. Soc. Am.*, **121**, 388–400 (2007).
[9] M. Vormann, M. Meis, V. Mellert and A. Schick, "A new approach for the evaluation of tonal noise (tonality)" in *Psychophysics, Physiology and Models of Hearing*, T. Dau, V. Hohmann and B. Kollmeier, Eds. (World Scientific 1999), pp. 109–112.
[10] DIN 45681: Acoustics: Determination of tonal components of noise and determination of a tone adjustment for the assessment of noise immission, Deutsches Institut fuer Normung e.V., Beuth Verlag GmbH, Berlin (2005).
[11] R. Bienvenue and M. Nobile, "Prominance ratio for noise spectra with discrete tones: A procedure based on Zwicker's critical band research," in *Proc. Internoise '91*, 53–56 (1991).
[12] E. Terhardt, G. Stoll and M. Seewann, "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *J. Acoust. Soc. Am.*, **71**, 669–688 (1982).
[13] H. Hansen, R. Weber and U. Letens, "Quantifying tonal phenomena in interior car sounds," in *Proc. Forum Acusticum 2005*, Budapest, Hungary (2005).
[14] H. David, *The Method of Paired Comparison* (Griffin, London, 1988).
[15] W. Ellermeier and G. Faulhammer, "Empirical evaluation of axioms fundamental to Steven's ratio-scaling approach: I. Loudness production," *Percept. Psychophys.*, **62**, 1505–1511 (2000).
[16] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, **39**, 324–345 (1952).
[17] F. Wickelmaier and C. Schmid, "A Matlab function to estimate choice model parameters from paired-comparison data," *Behav. Res. Methods Instrum. Comput.*, **36**, 29–40 (2004).
[18] A. Tversky, "Intransitivity of preferences," *Psychol. Rev.*, **76**, 31–48 (1969).
[19] M. G. Kendall and J. D. Gibbons, *Rank Correlation Methods* (Griffin, London, 1962).
[20] E. Zwicker and H. Fastl, *Psychoacoustics* (Springer, Berlin, Heidelberg, 1999).
[21] A. Turk and J. Sawusch, "The processing of duration and intensity cues to prominence," *J. Acoust. Soc. Am.*, **99**, 281–299 (1996).
[22] W. Garner, *The Processing of Information and Structure* (Lawrence Erlbaum Associates, Potomac, Maryland, 1974).
[23] F. Ashby and J. Townsend, "Varieties of perceptual independence," *Psychol. Rev.*, **93**, 154–179 (1986).
[24] B. C. J. Moore, B. R. Glasberg and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.*, **45**, 224–240 (1997).
[25] S. Kuwano and S. Namba, "Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness," *Psychol. Res.*, **47**, 27–37 (1985).
[26] S. Namba, S. Kuwano, T. Hatoh and M. Kato, "Assessment of musical performance by using the method of continuous judgment by selected description," *Music Percept.*, **8**, 251–276 (1991).

Hans Hansen received a degree in physics at Oldenburg University, Germany. He had a research stay at Prof. Kuwano's Lab at Osaka University, Japan. His main interests are acoustics and sound quality issues related to psychoacoustic parameters. He is currently working on his PhD, "Modeling tonality of technical and artificial sounds." Currently, he is an associate member of the "International Graduate School of Neurosensory Science and Systems" at Oldenburg University. He is a stipendiary of the German National Academic Foundation. He is currently with the "Institut de Recherche et Coordination Acoustique/Musique" in the Sound Design and Perception group.

Reinhard Weber studied physics at Goettingen University at the "Drittes Physikalisches Institut" in M.R. Schroeder's group. He received his PhD in 1976. Since then, his research field is psychoacoustics, the application thereof, and the perception of vibration. Since 1976 he is a researcher at the acoustics group of Prof. Mellert at Oldenburg University.